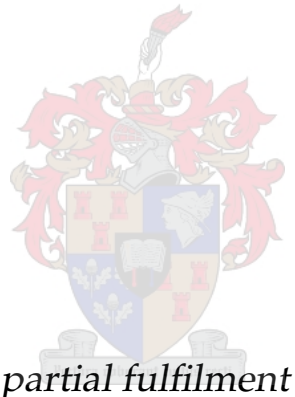
Towards a nuanced view of diagnostic test properties: an application to transfusion transmitted risk estimation

by

Jeremy Bingham



Thesis presented in partial fulfilment of the requirements for the degree of Master of Science (Mathematics) in the Faculty of Science at Stellenbosch University

Supervisor: Prof. Alex Welte

March 2021

Stellenbosch University https://scholar.sun.ac.za

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: 2021/02/17

Copyright © 2021 Stellenbosch University
All rights reserved.

Stellenbosch University https://scholar.sun.ac.za

Abstract

Towards a nuanced view of diagnostic test properties: an application to transfusion transmitted risk estimation

J. Bingham

Department of Mathematical Sciences,
University of Stellenbosch,
Private Bag X1, Matieland 7602, South Africa.

Thesis: MSc (Mathematics)

February 2021

Laboratory screening (rather than pathogen inactivation) is likely to remain, for the foreseeable future, the primary means of ensuring the safety of blood products from transfusion transmissible viruses such as Hepatitis B, Hepatitis C, and Human Immunodeficiency Virus (HIV). Depending on the tests used, there is generically some 'residual risk' of transfusion transmitted infection, as no test can guarantee detection of all potentially infectious material. Previously-described risk estimation approaches 1) mostly treat detectability and infectiousness as categories rather than continuously tunable; 2) disregard sources of variability and their correlation; and 3) are not generalizable to arbitrary detection biomarkers – making it difficult to generate estimates of residual risk without extensive programmatic monitoring.

We describe a broad framework for modelling test performance which incorporates hitherto neglected sources of variability in parameters governing the infectiousness and detectability of transfusion-transmissible pathogens. We utilise models based on this framework to demonstrate the relationship between test performance and residual risk for various assumptions about the

Stellenbosch University https://scholar.sun.ac.za

biomarker/infectiousness relationship, and illustrate how the same framework may be used to inform modelling efforts in related fields - such as infection dating and incidence estimation - which rely on realistic representations of test performance.

The key findings from our scenario modelling demonstrate: 1) Diminishing returns on increased screening sensitivity not evident in less flexible models; 2) increasing inter-subject variability in detectability and infectiousness leads to increasing residual risk in our general model, but lower risk estimates than in a previously described and widely used semi-mechanistic model. These effects are stronger when the average delay between infectiousness and detectability is short.

Planning blood product screening algorithms in light of simulations using our models can generate robust expectations of residual risk over a wide range of test performance and product risk levels. We outline when simpler models may be relied upon, and when additional nuance must be considered.

Stellenbosch University https://scholar.sun.ac.za

Uittreksel

Stappe tot ’n verfynde begrip van diagnostiese toetseienskappe: ’n toepassing op die beraming van bloedoortapping risiko

(“Towards a nuanced view of diagnostic test properties: an application to transfusion transmitted risk estimation”)

J. Bingham

Departement Wiskuudige Wetenskappe,
Universiteit van Stellenbosch,
Privaatsak X1, Matieland 7602, Suid Afrika.

Tesis: MSc (Mathematics)

Februarie 2021

Laboratoriumtoetsing (eerder as patogeeninaktivering) sal waarskynlik in die afsienbare toekoms die primêre manier bly om die veiligheid van bloedprodukte teen oordraagbare virusse soos hepatitis B en C en menslike immuniteitsgebreksvirus (MIV) te verseker. Afhangend van die toetse wat gebruikte word, is daar ’n ’oorblywende risiko’ van infeksie deur oortapping, aangesien geen toets die opsporing van alle moontlike aansteeklike materiaal kan waarborg nie. Voorheen beskryfde risiko-beramingsbenaderings 1) behandel opspoorbaarheid en aansteeklikheid meestal as kategorieë eerder as kontinue eienskappe; 2) bronne van variansie en die korrelasie daarvan word verontagsaam; en 3) kan nie veralgemeen word vir arbitrêre opsporing van biomerkers nie, wat dit moeilik maak om ramings van oorblywende risiko te genereer sonder uitgebreide programmatiese waarneming. Ons beskryf ’n breë raamwerk vir die modellering van die diagnostiese akkuraatheid van toetse. Dit sluit in

Stellenbosch University https://scholar.sun.ac.za

bronne van variansie in die parameters wat die aansteeklikheid en opspoorbaarheid van oortapping-oordraagbare patogene beheer. Ons gebruik modelle wat gebaseer is op hierdie raamwerk om die verband tussen toetsprestasie en oorblywende risiko vir verskillende aannames oor die biomerker / aansteeklikheidsverhouding aan te toon, en illustreer hoe dieselfde raamwerk gebruik kan word om modelleringspogings in verwante velde in te lig. Voorbeelde sluit in die beraming van infeksie datums en die beraming van infeksie insidensie, wat albei staatmaak op realistiese beramings van die akkuraatheid van diagnostiese toetse. Die belangrikste bevindings uit ons scenario-modellering toon: 1) Afnemende opbrengste met verhoogde sifting-sensitiwiteit, wat nie duidelik is in minder soepel modelle nie; 2) groter variansie tussen individue in opspoorbaarheid en aansteeklikheid lei tot groter oorblywende risiko in ons basis model, maar laer risiko beramings as in ’n voorheen beskryfde en wyd gebruikte semi-meganistiese model. Hierdie effekte is sterker wanneer die gemiddelde periode tussen aansteeklikheid en opspoorbaarheid kort is. Bloedprodukte-siftingsalgoritmes wat bepaal word deur simulasies met ons modelle, kan sterk verwagtinge van die residuele risiko oor ’n wye reeks toetsprestasies en produkrisikovlakke genereer. Ons gee ’n uiteensetting van wanneer eenvoudiger modelle gebruik kan word en wanneer addisionele nuanses in ag geneem moet word.

Stellenbosch University https://scholar.sun.ac.za

Acknowledgements

I am grateful to SACEMA and to the National Research Foundation for their financial and material support, without which this process would have been a great deal more difficult. I am grateful to my supervisor Prof. Alex Welte, whose sharp scientific mind and enthusiasm for sharing ideas have provided ample opportunities for learning, as well as to our collaborator, Dr. Eduard Grebe, who played a key role in conceptualising this work. I am grateful to the operations team and training coordinators at SACEMA for their outstanding and professional support. Without my friends, family, and colleagues I would surely not have prevailed - I am grateful for the many cheerful words of support and enthusiastic offers to proof-read my work. Finally, I am grateful to my parents, who taught me that it is okay to ask 'why', for the light of Love which they shine.

Stellenbosch University https://scholar.sun.ac.za

Contents

Stellenbosch University https://scholar.sun.ac.za

Stellenbosch University https://scholar.sun.ac.za

List of Figures

Stellenbosch University https://scholar.sun.ac.za

Stellenbosch University https://scholar.sun.ac.za

Chapter 1

Introduction

How often does a blood transfusion result in a chronic viral infection? Efforts to answer this question have been the subject of ongoing research since the mid 1980s [1]. Modern blood safety achievements have precluded the direct observation of transfusion-transmitted infection (TTI) risk due to the rarity of infection events. As a result, models are used to estimate the risk faced by transfusion recipients. Current methods rely on careful characterisation of the laboratory assays used to screen donated blood products before they are transfused [2], but may be limited by simplistic assumptions. In this work we set out to clearly describe the situations in which such simplistic characterisations are likely to be reliable and to explore key features omitted from established models of transfusion-transmitted infection risk.

Background

Blood transfusions are risky procedures when it comes to viral infections: transfused blood products invariably carry some risk of transfusion transmissible viruses such as hepatitis B (HBV), hepatitis C (HCV), and human immunodeficiency virus (HIV). Efforts to ensure a safe blood supply have driven crucial developments in the fields of virology, molecular biology, disease testing and, of course, transfusion medicine [1].

Waves of post-transfusion hepatitis during and after World War II led to the realisation that diseases could be transmitted (from donor to recipient) via blood transfusions [2]. Since this initial realisation, various efforts have been made

Stellenbosch University https://scholar.sun.ac.za

to estimate the risk faced by transfusion recipients. While modern safety measures have in many cases reduced this risk below the level where it may be directly measured, in the early days of TTI risk estimation, cases of the diseases of interest among both donors and transfusion recipients could be directly counted [2, 3, 4]. Efforts to quantify the risk of post-transfusion hepatitis (PTH) followed the discovery of the hepatitis B surface antigen (a 'biomarker' for post-transfusion hepatitis) and the subsequent development of the first viral assay for hepatitis [5, 6, 7]. Studies aimed at understanding demographic and behavioural risk factors for transmissible hepatitis measured rates of test-positivity for HBsAg in both recipients and donors. The results led to the banning of 'commercial' donations and a global shift towards volunteer-only blood collection [2]. In addition, by screening donated blood for HBsAg, blood banks were able to further reduce the remaining (residual) risk from donors who were not identified as being part of high-risk demographic or behavioural groups. Interestingly, the screening of donated blood products had a much smaller effect on risk than in donor eligibility requirements, though the same technology (viral screening assays) underpinned both measures [1, 2]. Despite the significant improvements to blood safety in the early 1980s, as-yet undiscovered and poorly understood pathogens would continue to threaten the safety of the global blood supply. The most shocking transfusion-transmitted pathogen, by far, was the human immunodeficiency virus (HIV).

During the early years of the HIV pandemic, it is estimated that hundreds of thousands of HIV infections were transmitted via transfusions, contributing significantly to the spread of the virus [2, 1]. The shockwaves that emanated from the realisation that HIV could be, and was being, transmitted via blood transfusions drove massive investment in HIV testing technology [8, 1]. Once the first reliable tests (first-generation antibody assays) became available, they were included in routine blood screening by virtually every blood bank which could afford it [1]. In conjunction with more stringent donor deferral practices (including deferring male donors who reported having sex with men) and donor self-deferral due to increased awareness, blood screening reduced the risk of transfusion-transmitted HIV (TT-HIV) by several orders of magnitude [2]. Nonetheless, some residual risk remained: infections were thought to occur mainly due to donations made by individuals who were recently infected

Stellenbosch University https://scholar.sun.ac.za

and still in the 'window period' for HIV (and thus would test negative) [9]. As tests became more advanced, researchers realised that measuring TT-HIV risk through direct observation was no longer feasible due to the rarity of these so-called 'breakthrough' infections. Efforts to overcome the challenges of risk estimation in the context of progressively safer blood products led to the development and introduction of the incidence per window period (IWP) model, paving the way for the current era of theoretical risk estimation [10, 11, 12].

Risk estimation in its most prevalent current form assumes the existence of a period of time (assumed to be the same for all recently infected blood donors) during which a person will definitely test negative on a routine screening algorithm, and whose blood will definitely cause an infection if it is transfused - the so-called 'infectious viraemic window period', or just 'window period'[10, 13]. This assumption has been supported by various successful predictions regarding the yield of updated screening algorithms (i.e. the number of new test-positive cases detected vs another algorithm). While the true risk to recipients is difficult to assess, and the small numbers of such infections which are detected preclude strong statistical inference, available evidence suggests that current methods tend to overestimate risk [2]. Recent approaches have produced more realistic estimates, and use well-justified assumptions, but are limited in that they apply only to a specific class of blood screening assay, and do not take into account inter-subject variability in the dynamics of early HIV infections [14, 15].

Aims

The aims of this work are to develop a generic, flexible approach to modelling the processes underlying post-screening TTI risk ('residual risk'), and to explore the implications for residual risk estimates. We also demonstrate the usefulness of this generic approach in the fields of incidence estimation and infection dating - fields in which modelling work relies on the accurate representation of test performance. To these ends, we address the following questions:

- What is the appropriate level of nuance to incorporate in a model of test performance?

Stellenbosch University https://scholar.sun.ac.za

- When are nuanced approaches necessary, and when may details be discarded in favour of simple summary parameters?
- What may be learned from a generic approach to modelling test performance?

Scope of this work

While many of the concepts presented in this work are relevant to a broad range of diseases and infections, differences in biology, available data, and relevant research aims rule out a completely general discussion of any depth. Specific test characteristics, interpretation rules for test results, and contexts in which tests are used may all influence results [16]. Where a closer focus is necessary we address issues related to the detection of 'early' HIV infections. HIV has received extensive research attention from a number of angles, yet remains in many ways mysterious [17, 18, 19]. In particular, much of the public health efforts surrounding the HIV pandemic, and accompanying research has had a strong focus on diagnostic tests [16, 1]. The result has been a wide range of increasingly accurate and precise HIV tests, along with a plethora of models utilising assumptions about these tests. While this necessarily reduces the generality of our implementation, we aim to describe the underlying motivations of the model in such a way that applying it to another disease would be as straightforward as possible. The steps we follow would be most easily generalised to other transfusion-transmitted viral infections (e.g. HBV, HCV).

The goal of this work is to explore the implications of a new approach to thinking about test performance. As such we will attempt to isolate the key details which must be considered for our purposes, and avoid overly-detailed modelling of underlying processes which are, and will likely remain, only roughly characterised.

Outline

The rest of this thesis is structured as follows. In Chapter 2 we introduce some necessary terminology and describe various types of tests for disease and infection. We discuss the basic test performance characteristics and the commonly

Stellenbosch University https://scholar.sun.ac.za

used methods for characterising test performance. We then discuss how the biology and epidemiology of early HIV infections affect test dynamics and require us to think beyond summary estimates of test performance characteristics. We describe a conceptual framework which incorporates these ideas and which may be applied to a wide range of problems. To demonstrate the framework's applicability, we then explore two adjacent areas of study in which a more nuanced view of in-situ test performance has already yielded important insights: incidence estimation and infection dating. In each case we offer a brief summary of key ideas and sketch how these may be thought of as special cases of the framework described.

Chapter 3 introduces the focal topic of this work: methods for estimating the residual risk of transfusion-transmitted infections. We provide an overview of the history of TTI risk awareness, prevention, and observation, highlighting the numerous challenges and successes encountered in the quest for safer blood supplies. We review in more detail the current prevailing risk estimation framework alongside some proposed alternatives, taking note of the benefits, structural limitations and rationale behind each approach. The incidence per window period model (IWP) [10, 11] and its successors [20, 13] have, since their introduction, largely defined the field of residual TTI risk modelling [2, 21]. As such, we discuss separately the risk estimation methods used prior to the IWP, the IWP with its successor, and alternative risk estimation models proposed since the IWP.

Through application of the general conceptual framework described in Chapter 2, Chapter 4 introduces a modification to the prevailing method for estimating risk, based on the idea that infectiousness and detectability of donor-derived material vary continuously over time for each donor. We discuss the benefits of this flexibility, as well as the challenges inherent in parameterising a more nuanced model. We then extend this model, as well as a more recent risk-estimation method, to incorporate the effects of variability between donors. Chapter 5 presents risk estimates using these extended models. The effects of various parametric and distributional assumptions are explored, and results are contextualised in terms of the underlying models.

Chapter 6 discusses the results, strengths and weaknesses in the proposed models, and suggests directions for future work. We compare our results to

Stellenbosch University https://scholar.sun.ac.za

those of published risk estimators, outlining the contexts in which the approach we introduce improves the performance of established methods. We highlight the value that the approach introduced in this work could bring to other modelling applications which rely on accurate representation of diagnostic test performance, and describe how planning blood product screening algorithms in light of the results from our model can lead to robust expectations of residual risk over a wide range of test performance and risk levels.

Stellenbosch University https://scholar.sun.ac.za

Chapter 2

Tests for disease and infection

2.1 Background

Tests for disease or infection may be categorised according to what they detect, as well as the purpose for which they are designed, benchmarked and approved.

Direct tests

Direct tests are designed to detect the presence of physical parts of pathogens themselves, such as nucleic acid, antigens, entire cells or multi-cellular structures. Nucleic acid amplification tests (NAT) have been developed to detect a variety of pathogens and are often considered 'gold-standard' tests, particularly for viruses [2]. Nucleic acid tests depend on carefully designed molecules that interact with specific sequences of nucleic acid (RNA or DNA) [22]. If the target RNA/DNA is present, the NAT procedure replicates target nucleic acids by many orders of magnitude, enabling direct observation (e.g via Gel Electrophoresis) [23].

'Quantitative' NAT testing - which outputs an estimate for the amount of target nucleic acid in the test sample - requires consistent levels of amplification so that final observed concentrations may be back-extrapolated to initial concentrations. Polymerase chain reaction (PCR) NAT assays approximately double the amount of target nucleic acid in the test well with every amplification cy-

Stellenbosch University https://scholar.sun.ac.za

cle, so that once the levels of nucleic acid are high enough to be measured, the original concentration may be inferred [23]. 'Qualitative' NAT assays - i.e. NAT assays which output a 'positive' or 'negative' result for each sample - have other advantages: for example, many transcription-mediated amplification (TMA) assays are able to detect multiple pathogens in a single test run [24]. This makes TMA assays a popular choice for the screening of donated blood products, typically against HIV and Hepatitis B/C. NAT testing may be extremely *analytically*[1] sensitive, and in some cases is thought to be able to detect a single viral nucleic acid strand present in the test well. In transfusion safety contexts, NAT tests are at times used in 'minipool' configuration, known as MP-NAT (rather than 'individual-donation' or ID-NAT). In MP-NAT, samples from multiple donors are pooled together before testing, effectively reducing the analytical sensitivity of the test.

Antigen testing (e.g. P24 antigen testing for HIV) uses chemicals designed to bond with viral material (antigens) normally targeted by (and which causes activation of) the immune system. Antigen testing is usually less analytically sensitive than nucleic acid testing because it does not include an amplification step, so antigen levels in the test sample must be high enough to give a clear result (i.e. appear above the background noise) [25, 24].

Culture tests involve inserting a sample of potentially-pathogenic material into a controlled environment conducive to pathogen growth (e.g. agar gel, host tissue, etc). After some time, if the pathogen is present, it will likely have reproduced to the extent that the pathogen or its effects may be directly/indirectly observed (e.g. via microscope analysis or spectroscopy). Culture tests are considered the gold-standard for a variety of bacterial infections (e.g. tuberculosis), but are less common for transfusion-transmitted viral infections [25, 26].

Microscopes may also be used to directly observe some types of pathogen without the need for culture-based 'amplification' [24]. This tends to be costly, particularly for smaller pathogens; as pathogens grow larger direct visual observation becomes cheaper and more effective. In some cases pathogens may even be observed directly by the naked eye (e.g. intestinal worms in fecal matter) [24].

[1] Analytical sensitivity refers to concentrations at which a test can detect a pathogen with specified levels of reliability - see Section 2.2.

Stellenbosch University https://scholar.sun.ac.za

Immune response

The human immune system is capable of responding to pathogenic threats in two broad manners. When dealing with an unrecognized pathogen, the immune system directs and produces generalised immune cells (e.g. macrophages) to attack the threat. While the generalised attack is underway, more specialised immune cells (e.g. T-helper cells, and later antibody-producing B cells) interact with the pathogen [27]. The presence of specialised cells for attacking a particular pathogen is therefore indicative that the pathogen has itself been present at some stage in the past. Antibody tests have been developed for a wide range of diseases and are the cheapest and most commonly used class of tests for HIV [2]. Antibody tests are often cheap to produce and provide information about well-controlled and past infections (which direct tests are typically unable to do). Antibody tests are generally less able than direct tests to detect very early infections since the immune system requires some time to mount a specialised response [25].

Outside of the laboratory

Likely the most frequently performed 'tests' for disease (and hence indirectly for infection) are not based on tissue samples, chemistry and carefully-controlled calibration, but on symptoms and case histories observed by healthcare practitioners, caregivers and even individuals 'testing' themselves. While we do not focus on these types of tests, the general approach we take could in theory be adapted to them.

Uses of tests

The other dimension in which tests can be categorised is the intended and/or formally approved *purpose* of the tests [25]. Medical professionals, health administrators and researchers use tests for disease and/or infection in a variety of ways. The three primary categories of test-use are clinical diagnostic, screening, and survey-related - these often overlap in the interest of efficiency and

Stellenbosch University https://scholar.sun.ac.za

ethics[2]. Clinical diagnostic tests are used to evaluate the condition of individuals with the general goal of informing personal health decisions. Screening tests are used to determine the condition of individuals and determine their suitability for a particular activity, such as moving between countries, competing as a professional athlete, or donating blood. Survey tests are used to generate data for research into the characteristics of the population and/or disease being studied. In both screening and survey tests it is common-practice to inform participants who test positive - though additional confirmation testing may be required. A particular purpose determines the characteristics of an 'ideal' test. The various fields in which tests are used have also driven the development of new tests which subsequently cross-over between use-cases. Notably, the development of currently available highly-sensitive HIV tests, as well as extremely low-cost rapid diagnostic tests (RDTs), was driven in part by the interplay between clinical practice and transfusion safety research [1].

2.2 Test performance characteristics: sensitivity and specificity

The most frequently cited performance metrics of a test for disease or infection are its diagnostic sensitivity and specificity, usually referred to simply as the test's 'sensitivity' and 'specificity', and defined as:

- Sensitivity: The probability that a 'true-positive' sample or subject will test positive.
- Specificity: The probability that a 'true-negative' sample or subject will test negative.

Tests for infection or disease must be formally characterised before they can be approved for use. While there are many features of prospective tests that must be specified or investigated, point-value estimates of sensitivity and specificity are widely used [28]. Definitions of test characteristics must in any practical

[2]For example, in many cases ethical standards require that study participants receive treatment - in such cases tests are used for both clinical diagnostic and survey-related purposes.

Stellenbosch University https://scholar.sun.ac.za

application be accompanied by a clear definition of 'true-positive'. This is usually determined by positivity on a gold-standard test, though in a few cases researchers have identified highly probable exposure times in subjects who subsequently tested positive, then considered the subject to be 'positive' subsequent to the presumed time of infectious exposure. Serial specimens gathered from such subjects are valuable for characterising early infection dynamics and measuring diagnostic delays [29, 30], since they can shed light on the distribution of 'eclipse' phases - periods following infectious exposure when gold-standard tests are unable to detect infection.

A less-frequently referenced property of laboratory tests is 'analytical sensitivity': this refers to the minimum quantity (or concentration) of target biomarker that a test can detect with some specified probability [31, 18, 15]. For example, the Procleix Ultrio Plus NAT assay has (according to its package insert) a diagnostic sensitivity for HIV of 100%, and analytical sensitivities for HIV of 94% at 20 'International Units' (IU) per ml and 59% at 6 IU per ml [32]. This example highlights another issue with characterising test performance - the lack of standardisation in measuring viral concentrations. So-called 'international units' are in fact estimated differently by different labs for each disease - in this case the publishers of the Procleix Ultrio Plus package insert (Griffols) have provided an estimate of 0.6 viral RNA copies per IU [32]. While the trend in HIV research is moving towards the less-ambiguous 'copies per ml', estimates using 'IU per ml' continue to be published [28].

Another performance property of tests is their 'diagnostic delay', referring to the delay following infectious exposure when a test can 'reliably' detect the infection. In this case 'reliable' detection may be taken to mean a detection probability close to the official diagnostic sensitivity of the test [24, 16, 20].

Given two tests for the same biomarker, the test with a higher analytical sensitivity will in general have a shorter diagnostic delay (i.e. be able to 'reliably' detect infections sooner). However, if the same test also has a higher rate of spurious failure, then it would in some contexts have a lower diagnostic sensitivity than the less analytically sensitive, but more reliable test.

Formal evaluations and characterisations of test performance usually address the complexity of characterising test performance by stating an achieved diagnostic (and at times analytical) sensitivity on a specified panel of samples

Stellenbosch University https://scholar.sun.ac.za

[29, 32, 28]. The sources and sizes of these panels vary from disease to disease. Blood samples used to benchmark HIV tests are generally frozen plasma or whole blood samples obtained from individuals whose time of infectious exposure has been established [29]. Ascertaining the time of infectious exposure requires that subjects have a single high-risk exposure in the period preceding their first positive test [16]. NAT tests are often published with an analytical sensitivity, obtained on standardised dilutions derived from material with known DNA/RNA concentrations [15]. While large-scale efforts to collect specimens from recently infected individuals has allowed some degree of overlap between the samples used in different test benchmarking efforts, high quality samples with multiple post-infection collection dates are difficult to obtain, and ethical obligations constrain the post-infection timeframe in which samples characteristic of untreated infection may be collected [29, 1]. Often only a subset of these test performance characteristics - diagnostic sensitivity/specificity, analytical sensitivity, and diagnostic delay - are reported by test manufacturers and approval bodies, though subsequent research has in many cases provided additional information [29, 30, 33, 34, 28, 26].

Test algorithms

A test algorithm is a set of procedures that specifies a combination of tests which are to be administered in a given context, along with a recipe for diagnostic interpretation including a precise final case definition [35, 28]. Algorithms may include multiple tests within one 'layer' (e.g. "all donations are screened using one antibody test and one NAT test") as well as conditional tests whose application depends on preceding test results (e.g. "patients are screened using rapid test 1; patients who test negative are tested using rapid test 2") [36]. For example, when HIV tests are applied in clinical settings, positive 'screening' results are often followed up by 'confirmatory' or 'supplemental' tests in order to decrease the probability that a true-negative patient will be identified as positive and undergo unnecessary treatment (i.e. in order to increase the net specificity) [37, 26].

Test algorithms may be described using the same types of performance metrics as their individual constituent tests. The number of specific assumptions

Stellenbosch University https://scholar.sun.ac.za

required in order to make clear statements about these properties increases with the complexity of the algorithm.

Consider a test algorithm consisting of two tests, with the following rule for diagnostic interpretation: the sample is classified as negative if either test returns a negative result. If the diagnostic specificities of the two tests are both 90%, and the diagnostic sensitivities of the tests are both 95%, the probability that a true-negative sample will be classified as negative - i.e. the specificity of the algorithm - is equal to one minus the probability that both tests return spurious positive results:

$$\begin{aligned} P(-_{\text{alg}}|\text{True}-) =& 1 - P(+_{\text{test A}}|\text{True}-)P(+_{\text{test B}}|+_{\text{test A}}, \text{True}-) \\ =& 1 - 0.1^2 = 99\%. \end{aligned}$$

Using the same classification rule, the probability that a true-positive sample will be classified as positive - i.e. the sensitivity of the algorithm - is equal to the probability that both constituent tests return correct positive results:

$$\begin{aligned} P(+_{\text{alg}}|\text{True}+) =& P(+_{\text{test A}}|\text{True}+)P(+_{\text{test B}}|+_{\text{test A}}, \text{True}+) \\ =& 0.95^2 = 90.25\%, \end{aligned}$$

i.e., less than the sensitivity of either constituent test.

On the other hand, the same testing procedure (i.e. every sample is tested using both tests) could be accompanied by a rule for diagnostic interpretation which classifies a sample as positive if *either* test returns a positive result. In this case the sensitivity of the algorithm will be higher than the sensitivity of either constituent test, since

$$\begin{aligned} P(+_{\text{alg}}|\text{True}+) =& 1 - P(-_{\text{test A}}|\text{True}+)P(-_{\text{test B}}|-_{\text{test A}}, \text{True}+) \\ =& 1 - 0.05^2 = 99.75\%. \end{aligned}$$

This classification rule gives the algorithm better sensitivity than the "classify sample as negative if at least one test returns a negative result" rule, but the specificity of the algorithm is reduced:

$$\begin{aligned} P(-_{\text{alg}}|\text{True}-) =& P(-_{\text{test A}}|\text{True}-)P(-_{\text{test B}}|-_{\text{test A}}, \text{True}-) \\ =& 0.9^2 = 81\%. \end{aligned}$$

Stellenbosch University https://scholar.sun.ac.za

Algorithms which condition the usage of a second 'confirmatory' or 'supplemental' test on the result of a primary test have similar trade-offs. If the confirmatory test is used for subjects who test positive on the primary test, and both results must be positive for the algorithm to give a positive diagnosis (i.e. one negative test causes a negative result from the algorithm), then the addition of the confirmatory test (relative to the single-test algorithm) improves the specificity of the algorithm at the cost of sensitivity.

The above examples assume that the imperfections of the tests in question are due to random effects. If the false negatives/positives from the tests are due instead to anomalies in the chemistry of the test samples (e.g. blood) from certain subjects, and the two tests have similar or identical chemical/physical mechanisms, then the probabilities of the tests returning false-negative or false-positive results are likely to be positively correlated. In the extreme case of perfectly correlated test results (i.e. $P(+_{\text{test B}}|+_{\text{test A}}) = 1$, $P(-_{\text{test B}}|-_{\text{test A}}) = 1$, and vice-versa), the algorithm's sensitivity and specificity would be equal to those of the constituent tests. If the tests are aimed at detecting the same biomarker, then correlation between spurious test results can only be reliably avoided with knowledge of the chemistry and production methods of the tests going into an algorithm, or data on the observed correlation between the tests [24]. Unfortunately, detailed information regarding the chemistry and manufacturing of tests is not always available to decision makers, nor are published characterisations of test performance (whether from the manufacturers or independent research) which include correlation data [32, 28, 25].

The costs associated with imperfect specificity and sensitivity depend on the context in which a test is used. False positives in a clinical setting may cause psychological trauma to patients and lead to unnecessary treatment, with the associated costs and potential side-effects. False negatives may lead to untreated infections and worsened health outcomes for patients, as well as increasing the potential for further transmission [24]. False positives in algorithms for screening donated blood products leads to wastage of usable blood, while false negatives may lead to infections in transfusion recipients [26].

The epidemiological context in which tests are performed can also play a significant role in their usefulness. This is usually captured with the idea of positive and negative predictive values (PPV and NPV respectively), which describe

Stellenbosch University https://scholar.sun.ac.za

the probabilities that positive or negative results reflect cases of true-positives or true-negatives . Given a test with fixed (imperfect) diagnostic sensitivity and specificity, and a negligibly short period of non-detectability following a subject's infectious exposure, the PPV and NPV will depend entirely on the prevalence [38]. The following table illustrates a situation in which a population of 1000 people, with a prevalence of 10%, is tested using a test with diagnostic sensitivity and specificity of 90%.

Table illustrating positive and negative predictive values of a test with diagnostic sensitivity and specificity of 90%, testing for a disease which is immediately detectable following infectious exposure, in a population of 1000 people with a true prevalence of 10%.

	True status	
Test outcome	Positive	Negative
Positive	90	90
Negative	10	810

The PPV of this combination of test and context is the proportion of of true positives among subjects who test positive:

$$\text{PPV} = \frac{\text{True positives with positive test results}}{\text{Positive test results}} = \frac{90}{90+90} = 50\%.$$

The NPV of the test in this context is

$$\text{NPV} = \frac{810}{10+810} = 98.8\%.$$

The PPV of a test increases with increasing prevalence because the imperfect specificity of the test causes a fixed proportion (in this simplified example) of the true-negatives to test positive. As such, the ratio of true-positives to test-positives decreases as the ratio of true-positives to true-negatives decreases. An in-depth discussion of positive and negative predictive values may be found in most introductory textbooks to biostatistics and epidemiology [39, 38].

While it is widely acknowledged that performance metrics such as PPV and NPV are dependent on context, what this may hide is the fact that a test's sensitivity and specificity *also* depend, though less drastically, on context. For example, the sensitivity of a test which has a long delay between infectious exposure

Stellenbosch University https://scholar.sun.ac.za

and reliable detection (which could be referred to as a test with a long 'window period') depends on the ratio of prevalence to 'recent' incidence (where 'recent' depends on the test's diagnostic delay). While this effect is likely small in a chronic infection, during an outbreak of a highly infectious disease the effective sensitivity of a test may depend significantly on its diagnostic delay. In surveillance contexts, screening tests may be chosen based on pre-survey estimates of prevalence and incidence, in order to bring the numbers of false-positives and false-negatives (which would ideally cancel each other out) nearer to each other [39, 38].

While distinctions may be drawn between different categories and uses of tests, the overarching concepts of test dynamics may be applied to any process aimed at disease detection. Indeed the less regulated and more variable nature of non-laboratory diagnosis illustrates well the unavoidable *variability* of test sensitivity and specificity. Somewhere between self-examination and procedural clinical diagnosis lie the pre-donation screening questionnaires used by blood transfusion services. These too may be thought of as tests, in the sense that they have some context-specific, and imperfect, sensitivity and specificity, each of which will have consequences that in turn depend on the rest of the test algorithm ('negative' questionnaire results lead to donation and laboratory screening, while 'positive' questionnaire results lead to deferral from donating). The results of the questionnaire will also vary with time for a particular subject as well as between subjects.

Even further past the realm of laboratory-calibrated test procedures are survey questionnaires and online data sources such as self-diagnosis websites (e.g. WebMD), Google Trends and social media platforms. While finding reference datasets from which to characterise the sensitivity and specificity of these data sources poses a significant challenge, they could for the purposes of surveillance all be thought of in the same framework, as imperfect tests with outcomes which will vary over time and between subjects.

Thresholds in test reactivity

The above discussion might lead one to think that test results are either 'positive' or 'negative'. In most cases however, this binarisation requires a rule

Stellenbosch University https://scholar.sun.ac.za

regarding how to classify some outcome measure (such as luminescence, absorption of light, etc.) from the physical test itself [24]. Classification rules vary in their complexity, but in their simplest form involve setting a threshold for test reactivity, or an analog of this, above/below which a test result is classified as positive/negative.

2.3 Acute and early HIV infection

The primary source of TT-HIV risk in blood transfusions is from recently-infected donors. The acute phase of an HIV infection[3] is characterised by:

- a period of viral establishment and relatively slow growth: this phase of an HIV infection is still poorly understood [40, 18], but the virus takes some time - thought to be on the order of 3 to 6 days [40] - to establish itself in the lymphoid tissue of the gut [43].
- once established in the lymphoid tissue of the gut, the virus begins a period of rapid growth, spreading to lymphoid tissue throughout the body and establishing reservoirs of latently infected cells. During this phase, which lasts on the order of 1-2 weeks [43, 40, 31] the viral RNA levels in plasma (the liquid component of whole blood) exhibit approximately exponential growth. This aspect of acute HIV infection is so widely agreed upon that it is a benchmark against which new models of acute HIV infection are validated [28]. During the early parts of the exponential growth phase, the immune system mounts a non-specialised ('innate') response to the infection. While this generalised defence is thought to be partially effective in reducing the reproductive capability of some virions, the increase in immune activity also leads to increased exposure (and hence destruction) of CD4 T-cells (the primary target of HIV), so that the overall effect may even increase viral growth [27, 17]. After some time (on the order of 20 days), the immune system begins to mount an adaptive response to the infection, targeting specific molecules on some of the HIV strains

[3] See [40, 41, 42, 17] for in-depth descriptions of the acute phase of HIV and more detailed modelling efforts.

Stellenbosch University https://scholar.sun.ac.za

present. Interesting to note is that adaptive immune responses take much longer to target founder virus strains, which are thought to be significantly more infectious than their generically variant descendants [40, 18]. It is also worth noting that HIV mutates rapidly, and genetic variants targeted by the first wave of adaptive immune response may resistant to it within 10-20 days [44].

- Due in part to the success of the adaptive immune responses, as well as the depletion of CD4 T-cells, viral RNA levels slow into sub-exponential growth before reaching a maximal 'peak viral load' and decreasing fairly rapidly to a minimum 'nadir' viral load, then settling at a roughly fixed 'set-point viral load'. The delay between infectious exposure and peak viraemia varies from subject to subject, and peak viraemia is thought to occur on the order of 25 days after infectious exposure [40, 42, 17].

2.4 A general framework for modelling test performance

Missing from the usual definitions of test sensitivity and specificity[4] is the accounting for the fact that, as time passes following a subject's infectious exposure to a pathogen, the probability that they will test positive, on a given test algorithm, changes [16, 14, 18]. In other words, a test algorithm's sensitivity is best represented as a continuous function of time. In the case of HIV screening algorithms, these changes are, effectively, monotonically increasing. Furthermore, the manner in which the probability of testing positive changes with time, varies from individual to individual.

That test performance changes with time-since-infection is a result of a more fundamental assertion: biomarker levels in a subject change over time. With the possible exception of the initial dose (or surgical intervention) biomarker levels do not jump instantaneously from one level to another but change smoothly over time. A model sufficiently general to capture every aspect of pathogens

[4] We focus on sensitivity for the remainder of this work; the concepts applied remain very similar.

Stellenbosch University https://scholar.sun.ac.za

which prey on humans would necessarily be extremely vague. These assumptions are chosen such that they apply well to HIV and other transfusion-transmitted viral infections. Our assumptions are as follows:

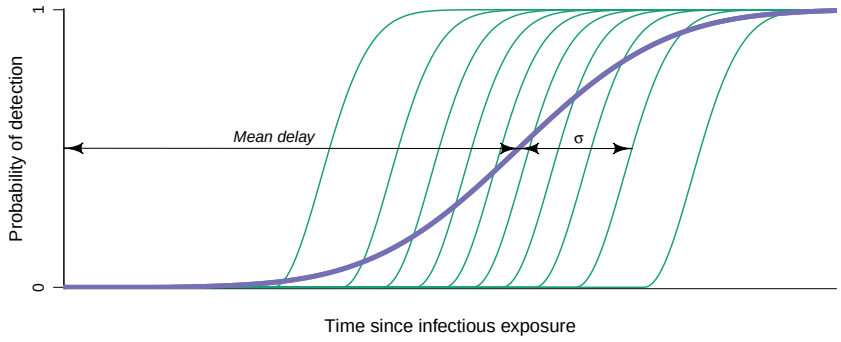


Figure 2.1: The time-varying detection probabilities of a hypothetical screening test. The thinner individual curves are generated using cumulative Weibull distributions, with normally-distributed positional shifts. The sigmoidal part of each individual curve illustrates the noisiness of measurements as each individual transitions from 'non-detectable' to 'detectable'. The shape of the thicker purple curve illustrates the average probability of detecting an infection at a particular time following infectious exposure. The shape of the mean detection probability curve depends primarily on the variability in the individual detection probability curves.

Firstly, at the time of a subject's first infectious exposure to a pathogen, the level of a given biomarker for the pathogen is indistinguishable from zero [16, 40]. This remains the case for some period following exposure.

Secondly, at some stage (since we are talking about an *infectious* exposure) biomarker levels will increase. This increase occurs smoothly relative to the scale of the host's body, and does not jump instantaneously from one value to another. As time passes, biomarker levels may fluctuate, plateau, or gradually increase/decrease, but will change in a smooth manner.

Thirdly, and as a result, the 'detectability' of a pathogen (i.e. the probability that a pathogen will be detected), using a given test algorithm (which may consist only of a single test) for a particular set of biomarkers, will change over time.

Stellenbosch University https://scholar.sun.ac.za

Initially, when biomarker levels are indistinguishable from zero, the probability that a pathogen will be detected is zero. As biomarker levels initially increase, the detectability of the pathogen also increases (see Figure 2.1).

Fourthly, for the purpose of residual risk estimation, we observe that the 'infectiousness' of blood products, from a donor who has recently experienced infectious exposure, changes as time passes. In most cases the infectiousness of blood products will be, effectively, zero at the time of infectious exposure and will (at least for some initial period) increase over time [15].

There are of course some exceptions to these rules. For example, in cases where repeat infections occur on short time scales, non-infectious matter from the pathogen of interest may still be circulating in the body at the time of a subsequent infection. Alternatively, if the dose at exposure is large and exposure occurs at the same point where test samples are taken or tests are performed (e.g. for a skin infection seeded by a severe exposure) initial biomarker levels might be significantly different from zero.

2.5 Incidence estimation

HIV incidence estimation is a key tool for monitoring the progression of the epidemic as well as for targeting and assessing public health interventions [45, 46]. While estimating the incidence of diseases with short life cycles, such as measles, is a *relatively* uncomplicated matter of counting new diagnoses, for chronic infections such as HIV the matter is much less straightforward. Longitudinal incidence studies which enroll and routinely test cohorts of HIV-negative participants are prohibitively expensive due to sample size requirements [46]. An alternative to longitudinal studies is the use of 'biomarker-based'[5] HIV incidence estimation [45, 46]. Biomarker-based HIV incidence estimation relies on specialised test procedures whose performance characteristics have been adjusted ('detuned') to make them much less analytically sensitive. In its original and most common form, this biomarker-based incidence estimation entails observation (i.e. testing) of a population using a standard diagnostic

[5]Note that while it is used in a more general sense in this work, the term 'biomarker' is used in the incidence estimation literature to refer to these detuned assays and accompanying analyses.

Stellenbosch University https://scholar.sun.ac.za

assay - typically an antibody test [47]. Individuals who test positive on the standard assay are then tested using the detuned tests, known as 'recency' assays, with those who test negative on the detuned tests being classified as 'recent'. This allows incidence to be estimated from a single cross-sectional survey, thus providing cheaper and more timely estimates of incidence [47].

A defining issue for the analytical work on biomarker-based HIV incidence estimation is the difficulty in making meaningful and accurate statements about test performance characteristics; while various biomarker-based methods have been developed, they often differ in the number of test performance parameters that must be estimated. Kassanjee et al. [46] showed that in general only two pieces of context-specific information regarding test performance characteristics are required in order to infer incidence from the combined results of standard and recency assays. These are a 'false recent rate' (FRR) specifying the rate at which 'non-recent' infections will be identified as recent under the test algorithm, and a 'mean duration of recent infection' (MDRI) [46]. The MDRI is defined by the integral $\int_0^T \mathrm{P}_{\mathrm{Recent}}(t)dt$, and is the mean time subjects spend in a 'recent' state between the moment in which they would likely test positive on the standard assay at hand, and some cutoff time T (correspondingly, the FRR is the rate at which subjects test as recent at or after the cutoff time T). The FRR and MDRI both depend on the choice for cutoff time T, as well as the threshold used to classify the results of the recency assay (i.e. as either recent or non-recent). In theory, a perfect recency assay would have an FRR of zero and a reasonably long MDRI, so that small samples could reliably be used to infer incidence (though the ideal MDRI would depend on the timescale for which incidence estimates were desired). Given a cutoff time T, the choice of threshold involves a tradeoff between an the MDRI and the FRR, as illustrated in Figure 2.2 below.

The field of incidence estimation has contributed in a number of ways to the understanding of HIV infections, test performance characteristics, and statistical methods for epidemiological surveys [48, 1]. One such contribution that is of interest to this work is the finding that models which assume that subjects transition in one direction through binary states of 'recent' (definitely testing negative on the recency assay) to 'non-recent', without any possibility of fluctuating test results (e.g. a single subject testing $+$'ve, $-$'ve, $+$'ve on consecutive

Stellenbosch University https://scholar.sun.ac.za

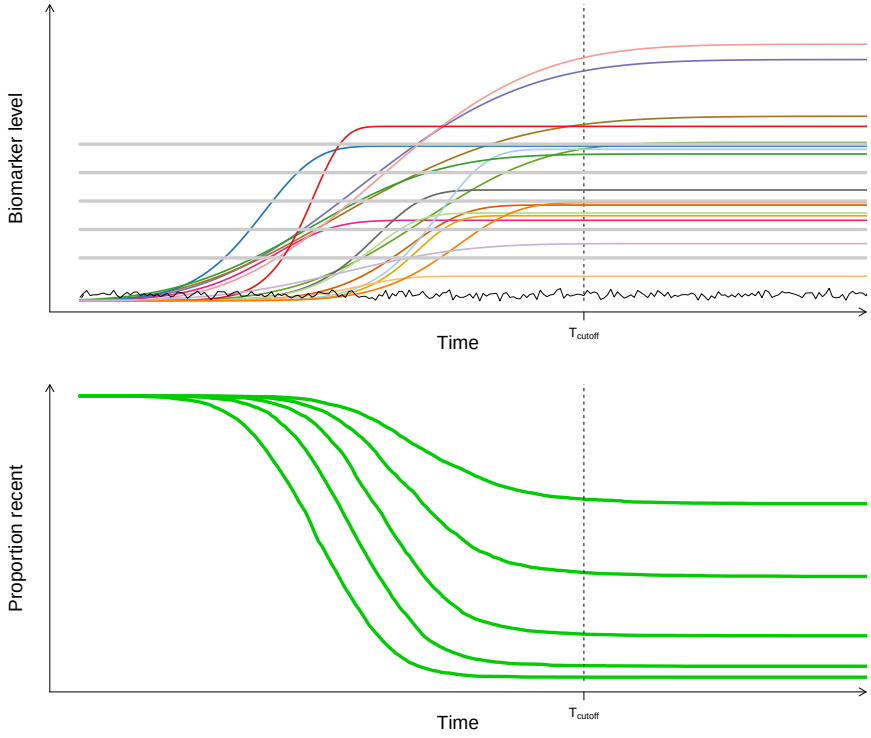


Figure 2.2: Recency assays require a choice of reactivity threshold, with implications for incidence estimation. Biomarker levels change in different manners for different subjects (colourful lines in the upper panel). Given a cutoff time T_{cutoff}, the choice of threshold (grey lines in upper panel) implies a mean duration of recent infection (MDRI), indicated by the green lines in the lower panel, and a false recent rate (FRR) [46].

tests), consistently underestimate the mean duration of recent infection (MDRI) [46].

2.6 Infection dating

Inferring the likely time of an individual's infectious exposure is frequently of interest to patients, caregivers and researchers alike [16]. From a research perspective, inferring likely times of infectious exposure is useful when collecting specimens for the characterisation of new diagnostic tests [30], estimating the frequency with which recently infected people present to donate blood [20], and understanding disease progression. Until recently, the only widely-used

Stellenbosch University https://scholar.sun.ac.za

infection dating system that utilised qualitative diagnostic test results was the "Fiebieg staging system" [13]. The Fiebieg system divides time-since-infection into a number of discrete stages during which subjects are presumed to test positive on certain tests, and negative on others. For example, a subject in the second "Fiebieg stage" tests positive on an RNA (NAT) assay and a P24 antigen assay, but does not yet test positive on an antibody assay [13]. This 'state', according to Fiebieg et al [13], lasts an average of 3.2 days, and begins an average of 13.9 days after infectious exposure. The Fiebieg system is limited in two key ways. Firstly, it is defined for a specific set of assays, many of which are no longer widely used or even available to purchase. Attempts have been made to create new versions of the system with modern assays, but as time passes and the number of available (and historical) assays increases, the usefulness of classifications such as "Fiebieg Stage 2" fades, and the amount of work required to fully characterise all the possible combinations of concordant/discordant test results increases rapidly [16]. Secondly, the Fiebieg system only accommodates discordant test results which take place at the same point in time; in practice, test samples may be collected on different days. In addition, the Fiebieg concept does not accommodate discordant test results for the same test, or for different tests with an 'unexpected' order of discordancy (i.e. a more sensitive test giving a negative result while less sensitive a test returns a positive result) [16, 13].

In a recently published article, we generalised the Fiebieg staging concept to incorporate arbitrary combinations of discordant and concordant test results, for tests taken at different times, without the need to specify the order in which subjects transition from one fixed 'stage' to another [16]. All that is required for the model to work is a time-varying 'detectability' curve for both tests, which specifies the probability that a subject tests positive, as a function of the delay between the time of infectious exposure and the time at which the test takes place. Since the time taken to transition from testing negative to testing positive is for most subject/test conditions short, compared to the inter-subject variation in delays between infectious exposure and testing positive, these detectability curves may be reasonably inferred from knowledge of a test's mean delay, combined with the variability in that delay. The likelihood that a subject was infected at a particular time may be calculated as the joint likelihood (in this case a product) of two test results, given the particular time of infection (Figures 2.3

Stellenbosch University https://scholar.sun.ac.za

and 2.4). A point estimate for the likely time of infection may be taken as the midpoint between the 50% detection probabilities for the two tests, or could in theory be calculated as an expectation value of the (normalised) likelihood function. The code base developed as part of this thesis (see Section 4.3) was used to implement the infection dating concepts described here, as well as to produce figures for the manuscript [16].

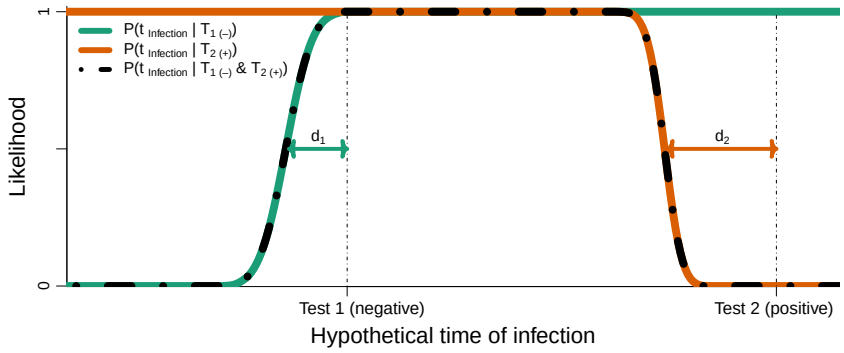


Figure 2.3: The likelihood of each hypothetical infection time (dashed black line) is given by the product of the independent infection time likelihoods given by each of the two simultaneous discordant test results (green and orange lines). d_1 and d_2 indicate the mean diagnostic delays to 50% detection probability of the two tests. The likelihoods for each hypothetical infection time, given the test results, forms a plateau roughly as wide as the gap between the two test dates, but not lying exactly between them.

In Figure 2.3, the black dashed curve (a soft-edged plateau) depicts the likelihood of each hypothetical time of infection leading to the observed (discordant) test results. A positive test result implies that the infection was not very recent, with positive results on less sensitive tests (i.e. tests with a longer mean diagnostic delay) providing more information than positive results on more sensitive tests (since less sensitive tests exclude a larger amount of recent time) [16]. On the other hand, increasing sensitivity makes negative results more informative. In theory (excluding test failure) the most informative discordant test results would arise when a less sensitive test shows a positive result while a more sensitive test shows a negative result - this is however highly unlikely to

Stellenbosch University https://scholar.sun.ac.za

occur. More likely is a subject showing different test results when tested multiple times using the same test on the same day (Figure 2.4) [16]. The information value of test results also depends on the shape of the tests' sensitivity curves - more variability across the population in delays to testing positive on a particular test results in a flatter sensitivity curve and less informative infection dating inferences. The shapes of the sensitivity curves are of less importance when the difference in mean diagnostic delays is large (as in Figure 2.3). On the other hand, when the difference in mean diagnostic delays is small, uncertainty in the timing of infection is driven primarily by the shapes of the test's sensitivity curves (Figure 2.4) [16]. See Section 5.1 for a related discussion, and [16] and [49] for a more complete description of this approach.

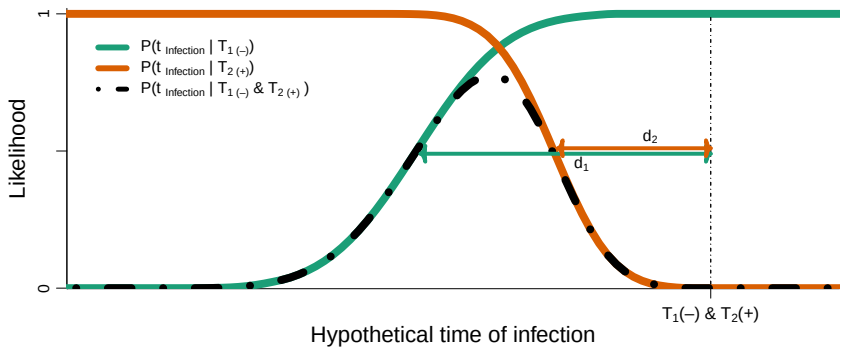


Figure 2.4: The likelihood of each hypothetical time of infectious exposure (dashed black line) is given by the product of the independent likelihoods given by each of the two non-simultaneous discordant test results (green and orange lines). d_1 and d_2 indicate the mean diagnostic delays to 50% detection probability of the two tests.

Stellenbosch University https://scholar.sun.ac.za

Chapter 3

A review of transfusion transmitted risk estimation methods

3.1 Transfusion-transmitted viral infection risk

Efforts to mitigate the risk of transfusion-transmitted infections (TTI's) in donated blood products have led to massive improvements in the safety of the global blood supply. Nevertheless, some risk remains - the primary source of risk for infections for which routine screening is in place comes from donations given after infectious exposure (when the donated material is potentially infectious) but before detection is probable [10, 2]. The amount of risk depends, along with epidemiological factors, on the safety procedures in place. These safety measures cost money and/or reduce the available supply of blood products, and questions naturally arise such as, "how much risk does each safety measure preclude?" or "how many infections might be prevented by improving safety procedures (or 'caused' by relaxing them)?". In order to address such questions, a number of risk estimation methods, also known as risk estimators, have been developed.

Risk estimators are useful for evaluating and comparing the public health implications of test algorithms, for example, when choosing among screening tests for various diseases given a limited budget, justifying additional expenditure on newly-available tests, or predicting the additional risk posed by alterations to donor eligibility criteria [1, 2, 50, 51].

Stellenbosch University https://scholar.sun.ac.za

Transfusion practice has evolved over the past five decades, and a full historical account would be impractical. Here we offer an introduction to the basic concepts in transfusion safety as they apply to the prevention of transfusion-transmitted HIV infections. The existence of chronic viral infections that could be transmitted via blood transfusions was first recognized during World War II, when post-transfusion hepatitis (PTH) was reported among US troops [1]. Studies to investigate and quantify the sources and extent of PTH risk began only in 1964 with a Boston-based retrospective study of PTH in transfusion recipients [52, 53, 1]. This study, which identified transfusion recipients with *symptomatic* PTH, found that the incidence of PTH was more than four times higher in recipients of blood from so-called 'commercial' donors (who were paid for their donations) than in recipients of blood from non-remunerated voluntary donors. The first prospective study of PTH was published in 1970 and showed a tenfold greater risk for recipients of commercial versus voluntary donations [54]. The study found that more than 50% of surviving open-heart surgery patients who received blood from commercial donors subsequently developed hepatitis [53]. These findings were instrumental in the United States Food and Drug Administration (FDA) and other regulatory agencies banning the use of commercial donors and legislating a volunteer-only blood donation system [55].

Another crucial step for blood safety made during this period was the identification of the Hepatitis B Surface Antigen (HBsAg) and methods for detecting its presence in blood samples [56]. Samples collected during the first prospective study of PTH were used to confirm the connection between the presence of HBsAg and clinical hepatitis. It had been estimated during this time that screening blood using first- or second- generation tests for HBsAg would eliminate approximately 25% of PTH [57, 58]; this was confirmed by a subsequent study [59], and from 1971 forward it became standard 'best-practice' to screen all donations for HBsAg [1]. As tests became more advanced, improved risk reduction via screening became possible. The persistence of PTH following these improvements led to the discovery of hepatitis C (initially called 'non-A, non-B hepatitis') in the early 1980s, and to the development of the first laboratory tests for hepatitis C [1, 54].

The blood safety successes of the 1970s and early 1980s were eclipsed, even as they were realised, by the appearance of 'transfusion-transmitted AIDS' (in fact

Stellenbosch University https://scholar.sun.ac.za

transfusion-transmitted HIV) in the early 1980s [1]. The shock of TT-AIDS led to a number of studies aimed at identifying risk factors for AIDS in potential donors [60]. Notably, the early studies on TT-AIDS involved creating massive repositories of specimens from blood donors. Such repositories proved invaluable both at the time and for decades afterwards [1]. Despite considerable effort and attention, it took some years before blood safety improvements were able to substantially slow the contribution of blood transfusions to the HIV/AIDS epidemic: retrospective studies estimated that in 1982 the risk of HIV infection per transfusion was $\approx 1.1\%$ in the United States [61]. As with post-transfusion hepatitis, the majority of transfusion-transmitted HIV (TT-HIV) risk was eliminated by modifying the donor pool to defer donors from groups shown to be at high risk of HIV infection [62]. The introduction of antibody screening (beginning in 1985) gradually eliminated most of the risk from donors who were not deferred [1, 2]. Subsequent advancements in screening tests further narrowed this residual risk and, in the current era, the risk of TT-HIV in many parts of the world has been reduced to miniscule levels [2]. A necessary side-effect of the improvements made to reduce the incidence of transfusion-transmitted HIV was that estimating risk directly from cohort studies, which already required large groups of transfusion recipients, ceased to be a viable means of risk measurement [2]. The primary tool remaining which could be used to observe new cases of transfusion-transmitted HIV, HBV, and HCV was the so-called 'lookback' study.

Lookback studies

Direct observational data on transfusion-transmitted infections is difficult to collect [2, 36, 63]. In order to estimate, based on observations of infection events, the per-incident-donor risk of infection, one would ideally like to know the total number of 'successful' donors (i.e. donors who passed the pre-donation screening questionnaire) who have experienced an infectious exposure, the infection status for all recipients of their post-exposure donations, and, in cases where both a donor and recipient test positive, the genetic profiles for both infections[1].

[1]Genetic information from infections in recipients can be compared with genetic information from infections in donors to identify which recipient infections were caused by transfused

Stellenbosch University https://scholar.sun.ac.za

The 'risk per incident donor' would then be:

$$\text{Risk}_{\text{per incident donor}} = \frac{\text{\# Infections caused by transfused material}}{\text{\# Incident donors}}.$$

Since the advent of modern transfusion medicine and growing awareness of TTI risk in the 1980s, donors who are known to be sick have been excluded, as have those with 'high-risk' lifestyle factors[2] such as injection drug use [26, 50]. As such, to observe the infection status of a donor, either the donor must be followed up and tested, or a sample of their blood must be stored at the time of donation and re-tested at a later date with a more-sensitive test than was used at the time of donation. Both of these data-collection strategies are expensive, and yield fairly uncertain data [26]; positive follow-up tests leave substantial uncertainty in the timing of infectious exposure, and hence in the status of the donor at the time of donation. In cases where stored samples are re-tested with a more sensitive test than was used at the time of donation, even the most analytically sensitive tests available may not detect every potentially infectious donation. These processes of donation, transfusion, and follow-up must be repeated many times in order to garnish an informative data set [29, 30]. In addition, many transfusion recipients die shortly after receiving a transfusion, ruling out direct observation of their potential TTI's [2, 36]. Rather than attempt to follow up and test every donor and recipient, researchers have opted for a more realistic data-collection strategy known broadly as 'lookbacks'. There are two types of lookback study: donor-triggered, and recipient triggered. A lookback study involves [64, 65, 66, 63]:

- Identifying an already-transfused donation which may have been infectious. Repeat donors (i.e. donors who have donated before, in contrast with first-time donors) who donate a test-positive blood sample can trigger the investigation, as can recipients who test positive and report the result to the blood service (usually via their healthcare provider).

- Locating and contacting the recipients and donors of the potentially infectious donation. If the lookback is triggered by a positive-testing donor,

material [36].

[2]The definition of 'high-risk' lifestyle factors has changed over time - see [50] for recent example.

Stellenbosch University https://scholar.sun.ac.za

confirmatory tests are performed on the donor; if the donor tests positive, the recipient is also tested. If the lookback is triggered by a positive-testing recipient, the donor is tested.

- If both donor and recipient test positive then case histories and, more recently, genetic analysis of the viral strains carried by each person, are used to investigate whether the recipient's infection was caused by the donation they received.

Names of transfusion recipients are often recorded in the administrative databases of transfusion services or healthcare providers administering transfusions [26, 36]. Based on these details, lookback administrators will attempt to contact recipients identified as being 'at risk' and request their participation in the study. None of the lookback steps are perfect, and some attempts fail; the success rate of the process varies widely over time and between countries. As many as 80% or as few as 15% of donors/recipient combinations might be contacted and tested in each lookback study [64, 65, 66, 63].

Once recipients of blood from an infected donor are identified, if they test positive for the pathogen in question it must be ascertained whether their infection was caused by the potentially-infectious donation. This may be done in a number of ways - in the first lookback studies, ascertainment of the source of infection was attempted by asking the donor and recipient detailed questions regarding potential non-transfusion exposure risks [1]. Transfusion was considered to be the cause of infection if no other likely infectious exposure could be identified, and if it were deemed 'possible' that the donor was already infected at the time of donation. In some modern studies, the genetics of the viral strains obtained from donor and recipient are compared to assess whether the infection was caused by the transfusion, or whether the recipient acquired the infection from another source [36]. As the world's transfusion safety has improved, cases of transfusion-transmitted infection have become rare, particularly in settings where rigorous lookback studies are performed and published [67]. Events worthy of lookback are sparse, and while methods for performing lookbacks have advanced (e.g. digital record keeping and affordable tools for genetic analysis), observed cases are few to be found [2, 36].

Stellenbosch University https://scholar.sun.ac.za

Despite their limitations, lookback procedures continue to be the most realistic method for observing transfusion-transmitted infections and assessing the accuracy of theoretical risk estimates. In countries with widely-deployed nucleic-acid testing, such as the USA and South Africa, theoretical calculations predict between 5 and 15 times as many transfusion-related HIV transmissions as lookback studies have observed [2, 36]. Likely sources for this discrepancy include:

- approximately half of blood transfusion recipients die soon after receiving their transfusion.
- recipients who die as a result of the health complications which necessitated transfusion may receive more transfused blood (since their injuries are more severe) than those who survive.
- some identified recipients (of potentially infectious donations) cannot be contacted, or refuse to take part in the study, and some donors may never return after providing an infectious donation.

While lookbacks remain the most realistic means for identifying transfusion transmission events, easier-to-observe quantities are often used to monitor TTI risk. For example, it is interesting to track the number of donors who test negative on one test, but positive on the other, as this provides some information about the timing of infectious exposures relative to donation [13, 20]. So-called 'NAT-yield' donations - donations which test positive on a NAT assay, but negative on all other screening tests (e.g. P24 antigen and antibody assays) - are a key data source for the prevailing risk-estimation method of the current era [20, 35] (see section 3.2 for details).

Lookback studies have played a critical role in the history of blood safety and virology. Evidence from lookbacks contributed to the discovery of the hepatitis C virus and to the introduction of NAT testing for HIV in many countries [1, 63]. They have also provided data which supports the idea that donations from an infected donor are, at least for some period of time following infectious exposure, neither completely infectious nor completely safe [68, 69].

Stellenbosch University https://scholar.sun.ac.za

Infectiousness

A key piece of information for estimating TT-HIV risk is the probability that a recipient of a blood product, from a donor who has experienced an infectious exposure prior to donating, will become infected [14]. We refer to this probability as the 'infectiousness' of the blood product.

The 'per-act' transmission risk for HIV has received considerable attention. While the focus has primarily been on sexual acts, for which many estimates exist [70, 71, 72], a number of attempts have been made to quantify the risk of infection occurring in recipients of transfused material derived from an HIV-infected donor [71]. These 'per-act' risk (i.e. infectiousness) estimates for transfusion are based on observations from the early years of the HIV pandemic, when large numbers of patients were transfused with blood from HIV infected donors who were in various stages of infection, and whose viral loads were controlled only by their own immune systems [71]. The results are and were staggering, showing very high average risk ($> 90\%$) for recipients of blood from infected donors, and helped drive substantial improvements to blood safety [71]. The high levels of infectiousness possible among infected donors highlight the need for continued vigilance and implementation of donor screening [2]. However, such estimates of infectiousness, which are from an era when established (non-recent) infections made up the majority of infection risk, have limited relevance when estimating TTI risk in modern blood banks, as the comparable group of donors would include those with long-standing infections, who are excluded from donating by a combination of pre-donation screening and antibody testing.

Studies estimating per-act HIV infection risk in areas other than transfusion (sexual acts, needle sharing, etc.) have shown that several factors affect the per-act transmission risk, including viral load, stage of infection, and the conditions to which the virus is exposed between leaving the host and potentially seeding a new infection [71, 70]. The occurrence of these factors has changed since the majority of data on the infectiousness of HIV via transfusion was gathered. For example, the cold-chain processes for blood post-donation have changed over time due to technological improvements and lower equipment costs [26].

Each type of donor-derived material has its own implications for residual TTI

Stellenbosch University https://scholar.sun.ac.za

risk, which in turn depend in knowable and un-knowable ways on the details of transfusion practice in a particular blood bank [37]. The most common types of donor-derived material are plasma, red blood cell (RBC) components, and platelet concentrates (PLT). In many blood banking systems, plasma is 'quarantined' in cold storage until a subsequent donation by the same donor is determined to be infection-negative, thus effectively eliminating TTI risk in plasma (from chronic diseases for which blood is screened) [26]. Red blood cell components, however, degrade when frozen and must be transfused within 42 days of collection [35]. While HIV does not infect red blood cells, when whole blood is fractionated into components, RBC components contain a small quantity of residual plasma. As a result, in contexts where plasma is quarantined, the primary source of transfusion-transmitted HIV risk is thought to stem from this small quantity (approximately 20ml) of plasma contained in each unit of red blood cell component [14].

In summary, the risk to transfusion recipients depends on both the prevalence and incidence of infections in successful donors (donors who pass the pre-donation screening), coupled with the performance characteristics of the screening algorithm being used, the type of material being transfused, and the conditions to which donated blood is exposed prior to transfusion.

3.2 The incidence per window period model

The first major purely-theoretical risk estimation method, and by far the most successful in terms of adoption, has been the 'incidence rate per window period model' (IWP) [1, 21, 35].

Since the widespread adoption of antibody screening and donor deferral policies in the late 1980s, TT-HIV risk in blood banks[3] has been primarily due to donations from recently-infected (or 'incident') donors, whose blood products may test negative on screening tests and subsequently cause infections in recipients [73, 21, 2]. The probability that an incident donor's donation will lead to an infection is related to the probability that their blood will test negative for HIV (and therefore be transfused), and the probability that their blood contains

[3]This applies to blood banks following the most basic modern safety guidelines. Blood banks which do not screen or test donors may still exist, but are not discussed here.

Stellenbosch University https://scholar.sun.ac.za

sufficient quantities of viable HIV virions to seed an infection in the transfusion recipient. In other words, the donor's infection must be simultaneously undetectable (or at least *undetected*) and infectious (at least, *cause an infection*) at the time of donation.

The IWP model assumes that donors transition between states of non-detectable and non-infectious, non-detectable but infectious (the 'infectious stage of the viraemic window period', or simply 'window period'[4]), and detectable (infectiousness is irrelevant if an infection is perfectly detectable) [11, 74, 10].

The model assumes that the 'window period' - in which the donor's blood tests negative on a given test algorithm, and is (potentially) infectious if transfused - begins at some time shortly after infectious exposure, and lasts for the mean diagnostic delay of the test being used [11, 73].

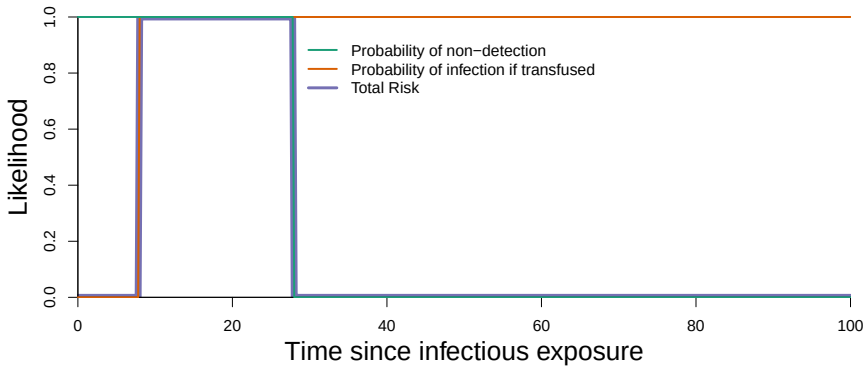


Figure 3.1: The so-called 'infectious window period' may be thought of in the general model as the product of two step functions, resulting in a perfectly flat-topped plateau. The 'window period' illustrated here represents the assumptions of the new incidence per window period model

At the time the model was proposed, 'incidence' could only be observed among repeat donors - donors who had donated before[5]. Donated material undergoes testing at every donation, so a test-negative donation followed by a test-positive

[4]Usage of term 'window period' varies somewhat with context.

[5]Donors who have donated before but who have not donated for a sufficiently long time are sometimes referred to as 'lapsed' donors.

Stellenbosch University https://scholar.sun.ac.za

donation indicates that infectious exposure must have taken place either between the donations, or during a short interval preceding the test-negative donation. First-time donors, on the other hand, test positive (or negative) at the time of donation, yielding no information[6] regarding whether their infection (if present) is new or long-standing. Positive-testing donors have, since the advent of antibody screening, been deferred from donating in future. Coupled with the fact that the potential window for exposure is much larger for first-time donors, this causes rates of test-positivity in first-time donors to be much higher than in repeat donors [20, 36].

In order to estimate the risk per incident donor, the size of the window period is divided by the average inter-donation interval of repeat donors:

$$\textbf{Risk}_{\textbf{per incident donor}} = \frac{\textbf{Window period}}{\textbf{Mean inter-donation interval}}. \tag{3.1}$$

Assuming that the donation rate of incident donors is unaffected by their infectious exposure, this can be interpreted as the probability that a donation from an incident donor will be given in the window period.

The probability that a donation takes place during the window period (which in the IWP is equal to the risk per incident donor) is then multiplied by the probability that a donation is from an incident donor, obtained by dividing the total number of incident donors by the total number of donations, to calculate the average risk, per transfused blood product, that a transfusion recipient faces. In equation 3.2 the total number of donations is multiplied by the average inter-donation interval in the denominator to produce a simpler expression.

$$\textbf{Risk} = \frac{\textbf{\# Seroconverting donors}}{\sum \textbf{Inter-donation intervals}} \times \textbf{Window period} \tag{3.2}$$

Two major concerns were voiced by the authors of the IWP at the time of publication:

1. The incidence and donation rate in first-time donors was not measurable, and had to be estimated in subsequent studies [7].

[6] This applies to the antibody-only HIV screening algorithms used at the time.

[7] Interestingly, the ratio of first-time to repeat donor incidence varied significantly across the world, even between regions with similar overall donor incidence [75]

Stellenbosch University https://scholar.sun.ac.za

2. The model assumes that donation rates are uniform, and independent of the timing of infectious exposure. It was suggested that donors might self-defer following risky exposure, and that people might knowingly donate blood following a risky exposure in order to obtain a free (high quality) test of their infection status.

The 'new Incidence per Window Period Model' (nIWP) was developed in response to these concerns and with the availability of new data from screening algorithms, which included NAT testing as well as antibody and (in some cases) P24 Antigen testing [20]. Based on the Fiebieg staging concept [13], in the nIWP, donors who test positive on NAT screening, but negative on antibody and/or antigen testing ('NAT-yield') are assumed to have recently experienced infectious exposure. The ability to identify donors who are HIV-positive but who have not yet 'seroconverted' (i.e. test negative on antibody tests) allows for more precise identification, and a new definition, of 'incident' donors. Data on donations made near to the time of infectious exposure are naturally relevant to residual risk calculations, and the NAT-yield approach to estimating incidence in the IWP confers two advantages:

- the NAT-yield approach allows observation of incidence in first-time donors, rather than crudely adjusting the incidence observed in repeat donors using estimates from other studies.
- the NAT-yield analysis requires weaker assumptions surrounding donation rates - there is a smaller time period (relative to infectious exposure) over which it is assumed that recently infected donors donate at a constant rate (for the original IWP, this period was the entire inter-donation interval).

The nIWP is still based on the concept of a 'window period' in which subjects are simultaneously infectious and undetectable, and assumes that infectiousness begins when a donor has an average viral load of 1 copy per 20ml of plasma - the viral load at which it becomes likely that at least one virion will be

Stellenbosch University https://scholar.sun.ac.za

present in the plasma portion of a single red blood cell component[8]. The size of the window period is calculated by back-extrapolating a model of exponential viral growth from the 50% detection threshold of a particular MP-NAT assay (Gen-Probe TMA) to the $\frac{1}{20}$ copies per ml assumed infectious limit. The rate of 'window period' donations is calculated based on the rate of donations which occupy a chosen Fiebieg stage:

$$\text{Rate}_{\text{WP donations}} = \frac{\text{Window period}}{\text{Fiebieg stage}} \times \text{Rate}_{\text{Donations in Fiebieg stage}},$$

where "Fiebieg stage" in the denominator indicates the mean duration of the chosen Fiebieg stage. Drawing on the Fiebieg staging concept introduces rich sources of data to the nIWP, but brings with it conceptual and practical limitations:

1. Fiebieg stages are defined for specific assays, none of which are widely used today, and few of which are even commercially available. While attempts have been made to define 'new Fiebieg stages' based on newer assays, the proliferation of tests makes characterising the numerous possible combinations unrealistic [16].

2. Individual donors are imagined to transition instantaneously between states of completely infectious and perfectly detectable.

3. All donors are assumed to experience Fiebieg stages (and window periods) of identical length - i.e. variability between donors and/or recipients is not considered.

The IWP[9] represented a groundbreaking shift in the way residual TTI risk was estimated [1]. The IWP has been widely adopted by blood safety researchers worldwide, and remains by and large the standard method with which to estimate residual TTI risk [35].

[8] While the authors of the initial IWP recognized that their model was (conservatively) estimating the risk that a 'potentially infectious' unit of blood would enter the blood supply (i.e. without quantifying the magnitude of the risk to the recipient), subsequent applications of the model have not always drawn this distinction [76, 77, 78]

[9] We refer hereinafter to the nIWP as simply the IWP; the differentiation is useful primarily for the purposes of historical review.

Stellenbosch University https://scholar.sun.ac.za

3.3 Semi mechanistic model (post-IWP)

Alternative approaches to the IWP have been proposed. Two primary directions have been taken which expand upon the IWP's approach. Bish et al. proposed risk estimation methods which leverage detailed manipulation of the mathematical expressions underlying incidence rate estimates and window period sizes [79, 80]. The approach, while interesting, relies on numerous unstated assumptions regarding the underlying processes, many of which are difficult or impossible to verify. Nonetheless, it highlights an interesting approach to the problem of designing blood screening algorithms, taken from an engineering perspective [81, 80]. The other main direction was developed by Weusten et al., who proposed a semi-mechanistic model for extending the IWP, with simple and transparent assumptions, and clear derivations [14]. The work of Weusten et al. has been generally accepted as a useful tool for blood safety modelling [36, 35]. We focus our application of the concepts described in Chapter 2 on the semi-mechanistic model proposed by Weusten et al.

In other cases, researchers have taken a completely different tack to estimating risk: for example, the analysis by Jayaraman et al. [82] suggests that WP risk is *not* the primary source of TT-HIV risk in sub-saharan Africa; they 'conservatively' assume that in countries for which information regarding blood screening practice is not 'available', blood is not screened at all. Despite such bold assumptions, some of the methods used to connect the assumptions with risk estimates appear reasonable. Nonetheless, the work seems unlikely to have any bearing on policy in the countries concerned (unless perhaps to prompt better sharing of blood screening policy). Indirectly estimating the risk in contexts without screening of donor-derived material is difficult to accomplish from a distance, since the sensitivity of pre-donation screening questionnaires depends on context, pre-donation screening procedures are less standardised than direct tests of donor-derived material, and incidence in the potential donor pool is hard to measure in such settings.

Stellenbosch University https://scholar.sun.ac.za

The semi-mechanistic model described by Weusten et al.

Weusten et al. have described a method for calculating the effective size of the IWP models' 'infectious viraemic window period' under NAT testing algorithms[10] [15]. In place of back-extrapolation to determine the magnitude of a strict 'gap' between an 'onset of infectiousness' and an 'onset of detectability', both infectiousness and detectability are modelled as functions of plasma viral load concentration, which is in turn modelled as a function of time [15, 14]. We follow here the derivations of Weusten et al. [14], expanding some of the steps for added clarity.

Starting at the beginning of viral doubling, the model uses a single viral load trajectory with exponential growth. Specifically, every donor's viral load is assumed to be given by $C(t) = C_0 2^{\frac{t}{\lambda}}$, where C_0 is the plasma viral load at the beginning of viral doubling, t is the time since the onset of viral doubling, and λ is the viral load doubling time. Since the model is designed to estimate risk in the context of NAT testing, it does not consider the dynamics of infection beyond the initial period of viral doubling. This appears to be reasonable since even less sensitive PCR NAT assays with large sample pools would typically achieve reliable detection rates within the viral doubling stage [14, 40, 18].

The model uses the viral load trajectory as an input to determine infection and detection probabilities (i.e. infectiousness and detectability). Infectiousness is modelled by assuming that virions in the transfused material have an average 'per-virion infectiousness' (p_V); that is, every virion which enters the recipient's body has a probability p_V of causing an infection, and a probability $1 - p_V$ of *not* causing an infection. The probability that *no* infection results from the transfusion of n virions is therefore $1 - (1 - p_V)^n$. Virions which are destroyed (or rendered non-viable for seeding infection) between donation and transfusion are still counted, and may be thought of as being incorporated into the value of p_V. One could equivalently increase p_V and consider only virions which survive through transfusion; these are mathematically equivalent in the model as described by Weusten et al. [14].

[10] In practice, blood which is screened for HIV using NAT assays is also screened for antibodies, since so-called 'elite controllers' as well as people on successful ART may present to donate and 'pass' the screening questionnaire.

Stellenbosch University https://scholar.sun.ac.za

The model further assumes that the number of virions transfused is a Poisson-distributed random variable, depending only on the volume V of plasma in the transfused component, and the concentration C of virions in the plasma, so that

$$P(n \text{ virions transfused}) = \frac{(C \cdot \mathrm{V})^n e^{-C \cdot \mathrm{V}}}{n!}.$$

The assumption of Poisson-distributed virion samples allows for arbitrarily large numbers of virions; this is considered an acceptable compromise for the sake of mathematical simplicity, since the probability of drawing unreasonably large numbers of virions vanishes quickly (e.g. with a plasma viral load of 1 copy/ml, the expected number of virions in one RBC unit is 20, and the probability of the component containing more than 42 virions is less than 0.001%). The 'infectiousness' of a plasma-containing blood product is therefore:

$$\begin{aligned}
&= \sum_{n=0}^{n=\infty} 1 - (1 - \mathrm{p_V})^n \frac{(C \cdot \mathrm{V})^n e^{-C \cdot \mathrm{V}}}{n!} \\
&= 1 - \sum_{n=0}^{n=\infty} \frac{(1 - \mathrm{p_V})^n (C \cdot \mathrm{V})^n e^{-C \cdot \mathrm{V}}}{n!} \\
&= 1 - \sum_{n=0}^{n=\infty} \frac{(1 - \mathrm{p_V})(C \cdot \mathrm{V})^n e^{-C \cdot \mathrm{V}} e^{-C \cdot \mathrm{V} \cdot \mathrm{p_V}}}{n! e^{-C \cdot \mathrm{V} \cdot \mathrm{p_V}}} \\
&= 1 - \sum_{n=0}^{n=\infty} \frac{(1 - \mathrm{p_V})(C \cdot \mathrm{V})^n e^{-C \cdot \mathrm{V} + C \cdot \mathrm{V} \cdot \mathrm{p_V}} \, e^{-C \cdot \mathrm{V} \cdot \mathrm{p_V}}}{n!} \\
&= 1 - e^{-C \cdot \mathrm{V} \cdot \mathrm{p_V}} \sum_{n=0}^{n=\infty} \frac{(1 - \mathrm{p_V})(C \cdot \mathrm{V})^n e^{-(1 - \mathrm{p_V}) C \cdot \mathrm{V}}}{n!} \\
&= 1 - e^{-C \cdot \mathrm{V} \cdot \mathrm{p_V}}. \qquad (3.3)
\end{aligned}$$

Each donation is accompanied by a set of samples used to screen for TTI's, including one or more samples used to screen for HIV. In the semi-mechanistic model described by Weusten et al., the probability that a sample will test positive on a single test is assumed to follow a probit function of the log of the viral load: $P(+) = \Phi(\mathrm{a} + \mathrm{b} \log(x))$, where a and b are parameters describing the test's properties, x is the average concentration of RNA particles in a subject's blood expressed in 'copies per aliquot of volume used in the test well', and Φ is a standard cumulative normal distribution function [14]. As mentioned in Section 2.2, a performance characteristic typically reported for NAT assays is the

Stellenbosch University https://scholar.sun.ac.za

analytical sensitivity - stated in terms of viral loads (concentrations) at which a test is able to detect specified percentages (e.g. 50% and 95%) of true-positive samples [32, 26]. If x_{50} is the viral concentration at which the test is able to detect 50% of positive samples, and x_{95} is the concentration at which the test detects 95% of positive samples, the parameters of the probit distribution can be inferred. Defining z to be such that $\Phi(z) = 0.95$ and noting that $\Phi(0) = 0.5$, it follows that

$$\begin{aligned} & a + b\log(x_{95}) = z \quad \text{and} \quad a = -b\log(x_{50}) \\ \Rightarrow\ & -b\log(x_{50}) + b\log(x_{95}) = x \\ \Rightarrow\ & b\log(\frac{x_{95}}{x_{50}}) = z \\ \Rightarrow\ & a = -\frac{z\log(x_{50})}{\log(\frac{x_{95}}{x_{50}})} \quad \text{and} \quad b = \frac{z}{\log(\frac{x_{95}}{x_{50}})}. \end{aligned}$$

Given a minipool size S_{pool} (with $S_{pool} = 1$ for ID-NAT), the probability that a donation from a recently infected donor will test negative at time t after the beginning of exponential growth is

$$P(-) = 1 - \Phi\left(\frac{z\log(\frac{n}{x_{50}})}{\log(\frac{x_{95}}{x_{50}})}\right) = 1 - \Phi\left(\frac{z\log(\frac{\chi C_0 2^{\frac{t}{\lambda}}}{S_{pool}x_{50}})}{\log(\frac{x_{95}}{x_{50}})}\right), \tag{3.4}$$

where χ is the number of RNA molecules per virion - two, in the case of HIV. This result was extended by Weusten et al. [14] to include multi-test screening procedures such as re-testing positive donations for confirmation ($P(-|k \text{ retests}) = P(-)^k$) or re-testing individual samples from pools which test positive ($P(-) = P(-)_{pool} \times P(-)_{ID}$); such procedures would in theory increase the total risk compared to discarding all samples linked to a positive-testing sample pool but may be necessary to reduce wastage, particularly when minipool testing is used [14, 26].

At any one time, the probability that a recently infected donor's blood will both test negative *and* cause an infection in the recipient is given by the product $P(\text{infectious}) \times P(\text{nondetection})$. The total number of 'risk-day equivalents' which each incident donor presents is then: $r_{days} = \int_{t_i}^{t_\infty} P(\text{infectious}) \times P(\text{detectable})dt$. The quantity of 'risk-day equivalents' naturally has units of

Stellenbosch University https://scholar.sun.ac.za

time, and corresponds in risk estimation to the same quantity as the length of the 'infectious viraemic window period'.

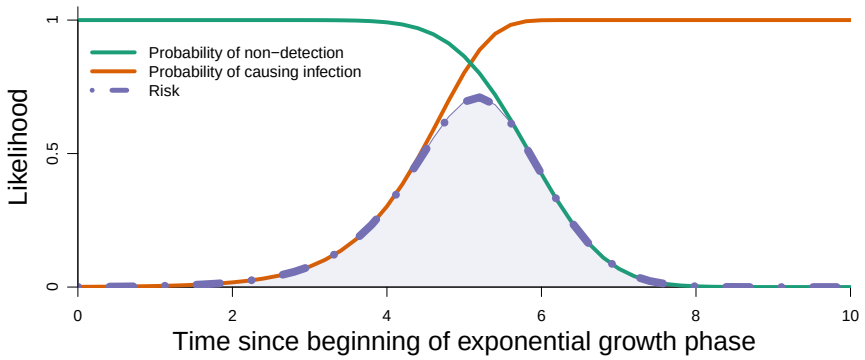


Figure 3.2: The probability that a red blood cell component from a recently-infected donor ultimately causes an infection in a transfusion recipient depends on the timing of their donation relative to the onset of exponential viral growth. This probability (a likelihood, when viewed as a function of time) is given by the joint probability (purple dotted line), that the screened sample will test negative on the screening algorithm used (green line) and will cause an infection *if transfused* (i.e. the infectiousness of the blood product, orange line), calculated at each point in time using the product of the individual probabilities. Calculation based on the model described by Weusten et al. (2011) [14], using a single Procleix Ultrio HIV assay in ID-NAT configuration [32].

Note that this model of nondetection probability does not explicitly consider the number of virions which appear in the test well, though it is possible for sensitive NAT tests to detect as little as a single virion. The probit analysis is used to accommodate the heterogeneity inherent in the sample collection process. At first glance, one might think to use a Poisson distribution - however, the small blood samples used for testing are not drawn from a large well-mixed container; virion concentrations in a subject's bloodstream may fluctuate over time, as the immune system responds to external stimuli, etc. Modelling this deviation from homogeneity explicitly would be challenging, and the probit analysis represents a middle ground between explicit modelling and simpler statistical models - it fits the available data better than a Poisson-distribution,

Stellenbosch University https://scholar.sun.ac.za

and can be thought of as combining the observed properties of the assay (i.e. the available data) with the inherent heterogeneity of the data collection process [83, 25, 84, 15].

The model described by Weusten et al. represents a solid step in the field of risk estimation, extending the incidence per window period model in a substantial and clear manner. The semi-mechanistic nature of the model balances pragmatic fitting to available data with available evidence regarding the details of acute HIV infections.

Characterising the semi-mechanistic model

In order to establish a baseline for comparison, we present a brief characterisation of how risk estimates using the semi-mechanistic model, proposed by Weusten et al., depend on various parametric assumptions.

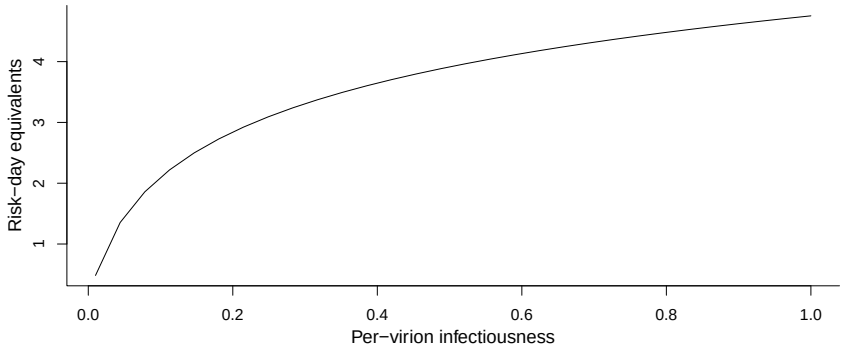


Figure 3.3: Risk day equivalents per incident donor, calculated using the semi-mechanistic model, for various values of 'per-virion infectiousness' p_V, under a screening algorithm consisting of a single Procleix Ultrio assay detecting HIV in ID-NAT configuration [14, 32].

Stellenbosch University https://scholar.sun.ac.za

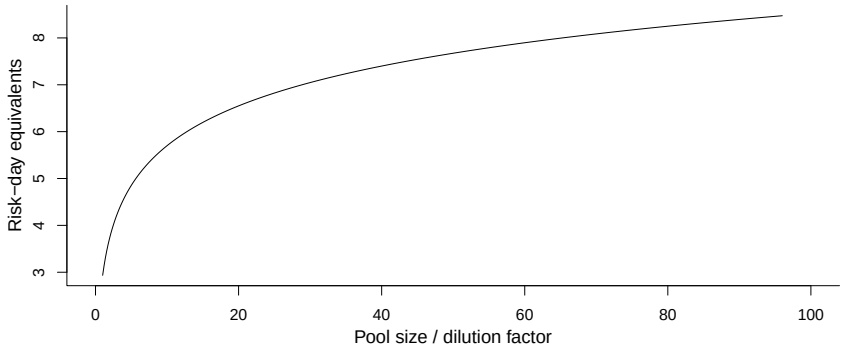


Figure 3.4: Risk day equivalents per incident donor, when testing pooled (or otherwise diluted) samples, according to the semi-mechanistic model. The 'pool size/dilution factor' may be thought of equivalently as the number of samples in a test pool, or as a reduction of the test's analytical sensitivity achieved by multiplying x_{50} and x_{95} by a factor equal to the pool size (see equations in previous section). The y-axis shows window period 'risk-day equivalents' under a screening algorithm consisting of a single Proxleix Ultrio assay detecting HIV in ID-NAT configuration [14, 32]. Decreasing sensitivity or increasing pool size causes diminishing increases in residual risk - e.g. switching from a pool size of 8 to a pool size of 16 would cause a larger increase in risk than switching from a pool size of 16 to a pool size of 32.

In the semi-mechanistic model, multiplying the analytical sensitivities (i.e. x_{50} and x_{95}) of the test by some factor is equivalent to multiplying the pool size by the same factor. This is because the detection probability depends on the viral load only in one term, in which it is divided by x_{50}. In the only other term containing x_{50} or x_{95}, these values are divided by one another. Therefore, multiplying x_{50} and x_{95} by some factor is equivalent to dividing (through dilution) the viral load by the same factor before putting it into the test well.

We also calculate the delay between 50% infectiousness and 50% detectability using various values of per-virion infectiousness (p_v) for the purpose of comparison with other models (Figure 3.6). Notice that the semi-mechanistic model predicts *some* residual risk even when the mean delay between infectiousness and detectability is zero or negative (when the standard window period calculations, e.g. from the IWP model, would predict zero risk). For larger values of

Stellenbosch University https://scholar.sun.ac.za

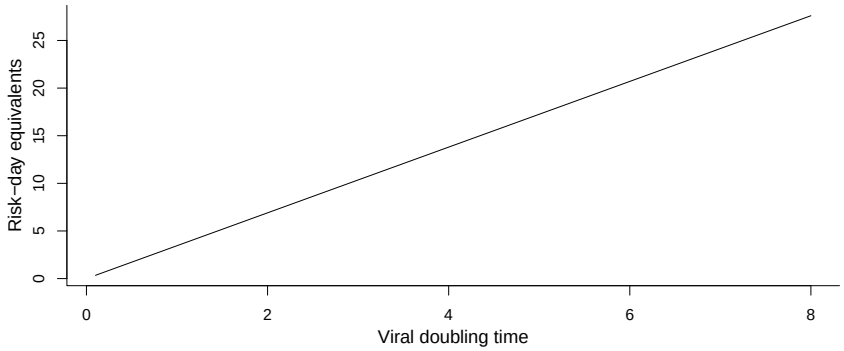


Figure 3.5: Risk estimates in the semi-mechanistic model vary linearly with viral doubling time (slope $= 3.4$). This occurs because the only model term in which time appears explicitly is divided by the doubling time, and the output of the model is an integral over time. Figure shows window period risk-day equivalents under a screening algorithm consisting of a single Procleix Ultrio assay detecting HIV in ID-NAT configuration [14, 32].

p_V however the models come into closer agreement.

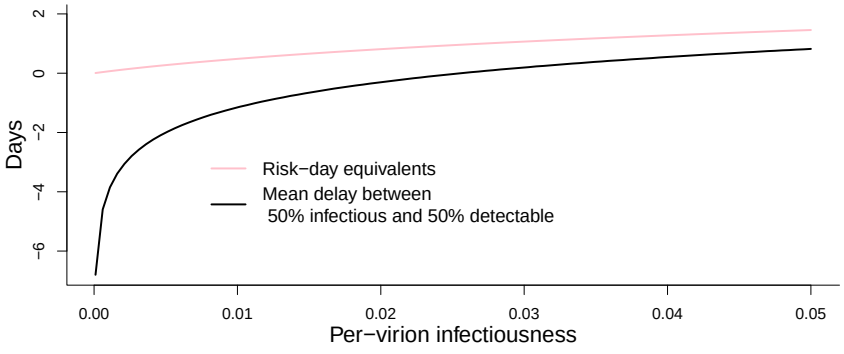


Figure 3.6: Risk day equivalents per incident donor, and mean delay between 50% infectiousness and 50% non-detectability, versus per-virion infectiousness according to the semi-mechanistic model. Figure shows window period risk-day equivalents under a screening algorithm consisting of a single Procleix Ultrio assay detecting HIV in ID-NAT configuration [14, 32].

Stellenbosch University https://scholar.sun.ac.za

Our implementation of the model described by Weusten et al. resulted in successful replication of their main results for 'no re-test' HIV screening (Table 3 in [14]) using the parameters they reported. We did however find one apparent error: it appears that the original authors mistakenly substituted into equation 3.4 the viral concentration in units of 'copies per ml' rather than in units of 'copies per aliquot of volume used in the test'. According to the package insert for the Procleix Ultrio assay [83], the standard volume of test material to be aliquoted is $500\mu l = 0.5$ ml. This means that concentration values in copies/ml must be halved before substitution. We replicated the authors' results by directly substituting the concentration, in copies/ml, into both equations 3.3 and 3.4 (i.e. using the incorrect units for equation 3.4). The use of incorrect units when substituting into equation 3.4 is equivalent to reducing the sensitivity (or increasing the pool size) by a factor of two. Accordingly we calculate, using the same parameters as reported by Weusten et al. (but corrected for the volume of material aliquoted), a WP risk-days equivalent under ID-NAT of 3.74 days (which is the same value as Weusten et al. reported for minipool testing with a pool size of two), rather than the 2.93 days calculated by Weusten et al. [14].

In addition, it is not clear, based on the references given by Weusten et al., where the estimates used for x_{50} and x_{95} are obtained (they use $x_{50} = 2.7$ and $x_{95} = 18.4$ copies/ml). The package insert for Procleix Ultrio HIV-1 (the HIV assay referenced by Weusten et al. [14]) claims 50% and 95% limits of detection of 13.9 copies/ml and 37.7 copies/ml respectively. The two independent studies examining the performance of Procleix Ultrio for detecting HIV (as well as HBV and HCV), which Weusten et al. cite as their sources, report central estimates for x_{50} and x_{95} of 5.31 IU/ml $\approx$ 3.19 copies/ml and 32.9 IU/ml $\approx$ 19.74 copies/ml [85], and 6.1 IU/ml $\approx$ 3.66 copies/ml and 42.2 IU/ml $\approx$ 25.3 copies/ml. Several other values are also reported (reflecting estimates made other panels of preserved samples), but none of these match up with the values reported by Weusten et al.

Stellenbosch University https://scholar.sun.ac.za

Donation timing

After incidence, which is *relatively* straightforward to measure, another 'ingredient' to the incidence per window period model and its derivatives (including those described in this work) is the effect of infectious exposure on donors' rates of donation [20, 2]. The standard assumption has remained that donation rates are constant; in particular, that donation rates are constant relative to infectious exposure [35]. However, the fact that antibody testing eliminates all but a small window of infection risk for each donor who donates following infectious exposure means that a even a short period of self-deferral practiced by a potential donor (following their infectious exposure) would drastically reduce their contribution to the overall risk faced by transfusion recipients. It is reasonable to think that many repeat, seroconverting donors (who regularly pass the pre-donation screening questionnaire) may be infected due to periods of unusually risky exposure (rather than generally risky lifestyles), which might lead them to self-defer (or be deferred during questionnaire/interview screening) in the periods during which they would contribute most to the overall risk [2].

Our exploration thus far suggests two possible sources of risk over-estimation: 'conservative' assumptions regarding per-virion infectiousness, and the assumption that donation timing is unaffected by risky exposure (so that incident donors are equally likely to donate at any time point relative to infectious exposure). Two mechanisms have been suggested through which donation rates might be modified by risky exposure. The first explanation posits that non-remunerated voluntary donors are generally well-meaning people, and as such are likely to 'self-defer'[11] from donating blood after periods of risky exposure (for example, after having unprotected sex with new partners). The second explanation is that the desire to access a free, high quality HIV test following periods of risky exposure leads people to present themselves for donation, and attempt to pass the pre-donation screening, at the time when they present the highest level of risk to potential transfusion recipients [86]. A number of published works (doubtless alongside many unpublished analyses) have attempted to identify potential

[11] 'Self-deferral' in this contexts includes any mechanism that would prevent a potential donor from actually giving blood, including answering survey questions honestly in such a way that they are deferred from donation, as well as simply choosing not to attend a blood donation opportunity.

Stellenbosch University https://scholar.sun.ac.za

deviations from constant donation rates among seroconverting donors, but the available evidence to date appears to support the idea that donation timing is not affected by high-risk exposures [86, 36].

Stellenbosch University https://scholar.sun.ac.za

Chapter 4

A general framework for estimating transmission risk

The general modelling framework introduced in Chapter 2 encompasses previously described methods, such as those in the IWP and the semi-mechanistic model, for calculating the effective size of the 'infectious viraemic window period' [20, 14]. In this section we describe some implementations of the general modelling framework. Focusing on the notion of *variability* among donors, recipients, viral strains, conditions to which donated blood is exposed, and transfusion procedures, we apply the general modelling framework to calculate the effective size of the 'infectious viraemic window period'. Following the language of Weusten et al., we refer to this quantity, which has units of 'days', as the 'risk per incident donor', measured in 'risk-day equivalents'[1]. The specific assumptions used in the present work are described and a code base is introduced for implementing flexible models of test performance.

4.1 Intra- and inter-subject variability of infectiousness and detection probability

Many details of 'the truth' behind the assumptions in TTI risk models are destined to remain fairly uncertain over the foreseeable future. Residual risk for

[1] 'Overall' risk might instinctively be thought of as the per-transfusion risk to the recipient - this is directly proportional to the quantity which we refer to as 'overall risk'.

Stellenbosch University https://scholar.sun.ac.za

TTI's depends on a combination of biology, epidemiological and social context, and test properties [2]. Nonetheless, we wish to establish the contexts in which certain properties of the general model have a significant impact on risk estimates. To begin, we review our general model's assumptions in slightly more detail.

Firstly, at the time of a hypothetical subject's infectious exposure (also referred to as the 'time of infection'), the probability of a test algorithm detecting the infection (the 'detectability' of the infection under the test algorithm) is zero [16]. After some time has elapsed, the probability of detection for the subject (as well as the average probability across all hypothetical subjects with the same time of infectious exposure) will increase, following the increase in target biomarker levels, until it is reaches one[2] [16]. In other words, the probability that the algorithm will *fail* to detect the new infection will, after some time has elapsed, decrease from one to zero (see Figure 4.1).

Similarly, at the time of infectious exposure, a subject's blood has effectively zero probability of causing an infection if it is transfused. This is considered reasonable because the initial phases of HIV infection are thought to occur in a localised fashion, with no viable virions or infected cells circulating in the bloodstream[3] [19, 88, 40]. Furthermore, it is widely believed that the infectiousness, via transfusion, of a subject's blood - specifically, one unit of red blood cell (RBC) component - increases during initial infection until it approaches one (see Figure 4.2 below) [89, 71]. The true maximum infectiousness of HIV-infected blood is difficult to estimate and may depend on a variety of factors including the type of blood product being transfused, the handling of blood products (e.g. centrifuging, cooling/warming and transportation) between donation and transfusion, and the particular characteristics of the virus, donor and recipient, but evidence from the early stages of the HIV pandemic suggest that it has a mean of at least 95% [71]. In this work it is assumed that infectiousness may come arbitrarily close to one, though this assumption could be easily relaxed. The trend in transfusion risk modelling has been to 'conservatively' estimate

[2]The diagnostic sensitivities reported for approved screening tests are typically very close to 1 [87, 37].

[3]A recently infected transfusion *recipient* might in theory be an exception to this, but they would not be donating blood immediately after transfusion.

Stellenbosch University https://scholar.sun.ac.za

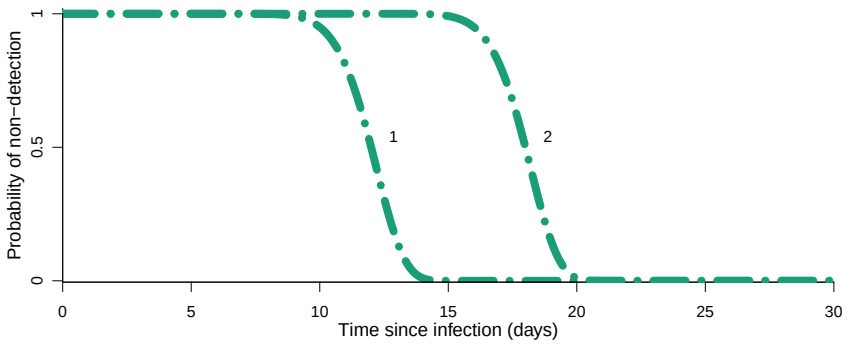


Figure 4.1: The probability that a subject who recently experienced an infectious exposure will test negative changes with time. In this figure, two individuals have been assigned characteristic delays to 'detectability' (and corresponding 'non-detection probability'). The sigmoidal piece of each curve illustrates how noisy the measurement is as they approach detectability (i.e. the noisiness in the process of going from non-detectable to detectable), while the horizontal position reflects the delay (e.g. to 50% detectability). The two hypothetical subjects pictured differ only in their delays, but in general different individuals would also have differently shaped transitions from non-detectable to detectable.

that every virion is completely infectious, and that the probability of a recipient being infected is equal to the probability that the blood product they receive contains at least one virion (an exception being the work of Weusten et al.) [14, 21]. As mentioned in Chapter 3, the authors of the IWP recognised that they were estimating the risk of a 'potentially infectious' (rather than definitely infectious) donation entering the blood supply. In that sense, this work (and the work of Weusten et al.), is estimating a new quantity - the risk of infection per incident donor - rather than the old 'risk of potentially infectious blood'.

The risk faced by each recipient of blood from a hypothetical donor who donates following infectious exposure is given by the product of the donor's time-varying 'infectiousness' and 'non-detection probability' (see Figure 4.3 below).

The final characteristic of the general model is that of donor, virus and recipient variability. Biological systems are full of variability, both random (or unexplained) and structured. The immune systems of transfusion recipients, for example, are likely to vary widely, due in part to long term factors such

Stellenbosch University https://scholar.sun.ac.za

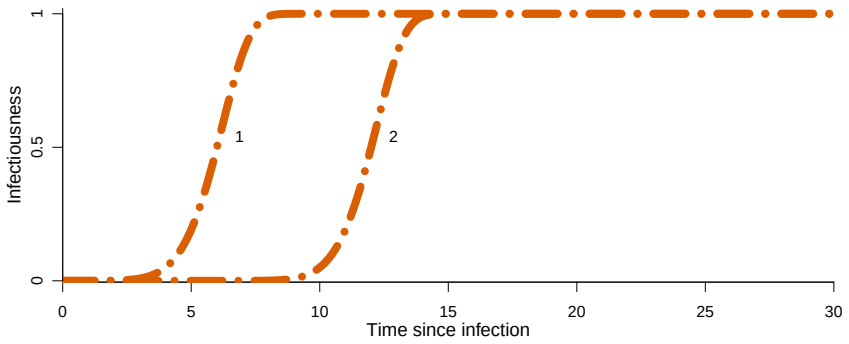


Figure 4.2: The infectiousness of blood products, from a donor who has recently experienced infectious exposure, changes with time. In this figure, two individuals have been assigned characteristic delays to maximal infectiousness (in this case maximal infectiousness equals 1). The sigmoidal piece of each curve illustrates the noisiness in the process of going from non-infectious to infectious, while the horizontal position reflects the delay. The two hypothetical subjects pictured differ only in their delays, but in general different individuals would also have differently shaped transitions from non-infectious to maximally infectious.

as age and genetics, as well as acute factors (e.g. factors related to the recipient's need for a transfusion) [27]; some recipients may need transfusions due to conditions which affect the immune system drastically, while for others the immune system is functioning as usual. Pathogens too vary - HIV is renowned for its genetic variability [44]. For example, natural selection often enforces trade-offs between pathogen virulence and transmission opportunities allowed by host longevity. For HIV in particular, there is evidence to suggest that founder strains - the genetic variants which seed infections - are much less successful at reproducing within hosts, so that once infections are well-established, the most prevalent strains are much less effective than the founder in terms of between-host transmission [90, 91, 92]. This in turn suggests that there might be correlation between factors influencing within-host reproduction and factors influencing between-host transmission (i.e. a correlation between growth rate and infectiousness).

There is probably at least some positive correlation between the times (fol-

Stellenbosch University https://scholar.sun.ac.za

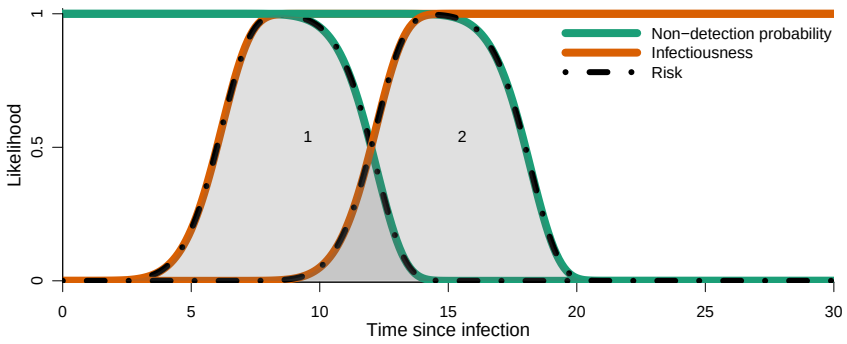


Figure 4.3: The risk that donated blood from a recently-infected subject will both test negative, and cause an infection in a hypothetical recipient, changes with time. In this figure, two individuals have been assigned characteristic delays to detectability and maximal infectiousness. The likelihood of simultaneous non-detection and subsequent transmission, for each point in time, is given by the product of the individual likelihoods. The two hypothetical subjects pictured differ only in their delays, but in general different individuals would also have differently shaped transitions from non-infectious to maximally infectious and non-detectable to maximally detectable.

lowing infectious exposure) at which subjects begin to be detectable on a certain assay, and the times at which they begin to be infectious [16]. For example, those who become detectable early (e.g. reaching 50% detectability early relative to a population average) on a particular NAT assay are likely to reach 50% infectiousness earlier than average, since pathogen levels are a key determinant of infectiousness, and NAT assays measure the pathogen directly [18]. On the other hand, one might expect slightly weaker (though still positive) correlation between the timing of 50% detectability on an antibody assay and the timing of 50% infectiousness. Random factors, such as the quantity of plasma left in a particular unit of RBC component after fractionation or the exact volume of sample material aliquoted into a test well, will also tend to reduce the strength of correlation. Given the multitude of difficult-to-observe factors which might effect the strength and direction of correlation, we explore the effects that different strengths and directions of correlation have on the overall residual risk.

Stellenbosch University https://scholar.sun.ac.za

Modern HIV tests used for transfusion screening achieve for the most part[4] extremely high diagnostic sensitivity levels and, when used in sensitivity-prioritising multi-test screening algorithms [37, 2], eliminate virtually all the risk from long-infected donors[5]. In other words, the single significant source of TT-HIV risk, in settings following recommended blood safety procedures, is recently-infected donors whose test samples contain inadequate biomarker levels for reliable detection (i.e. who would be described as being in the test algorithm's window period) [20, 2].

4.2 Generic residual risk model specification

Time-varying infectiousness and detectability curves are represented using cumulative Weibull distributions. Specifically, the likelihood of a positive test result at time t after infectious exposure is given by:

$$P(\mathrm{T}_+) = \mathrm{Sens}_{\mathrm{diag}} \times (1 - e^{-(\frac{t-\epsilon}{\gamma})^k}) \text{ if } t > \epsilon,$$

where $\mathrm{Sens}_{\mathrm{diag}}$ is the diagnostic sensitivity of the test in question (henceforth assumed to be 1), ϵ is the delay between exposure and the first increase in the probability of testing positive[6], k is the 'shape' parameter, and γ is the 'scale' parameter. The Weibull distribution is one of a handful of commonly used functions for modelling within-host disease dynamics with few parameters or little information regarding the mechanism of disease [94, 95]. In line with the aims of this work, the cumulative Weibull distribution is flexible, satisfies the criteria described in Section 4.1 for infectiousness and detectability (transitioning smoothly from some baseline value to some maximum value) and has relatively few parameters (two, excluding positional shift) [96, 97, 22].

[4] Cost constraints may limit the screening algorithm to tests with lower-than-recommended sensitivities due to inadequate laboratory controls and/or the use of rapid diagnostic tests.

[5] A potential exception is infected donors on long-term ART treatment - e.g. see [93, 50].

[6] The probability of testing positive when t $<= \epsilon$ is equal to the diagnostic specificity of the test in question.

Stellenbosch University https://scholar.sun.ac.za

Delay between 5% and 95% on a Weibull curve

'Individual' curves - that is, the curves representing hypothetical subjects - are parameterised such that they align with available data regarding the time taken to transition from non-detectable to highly-detectable, and from non-infectious to highly infectious[7]. If an individual has a 5% probability of testing positive at time t = a, 95% probability of testing positive at time t = b, and we specify the 'shape' parameter k, then the scale parameter may be calculated:

$$\gamma = \frac{a - b}{\sqrt[k]{\ln(0.05)} - \sqrt[k]{\ln(0.95)}}.$$

Fiebieg et al. report 'highly consistent' patterns of test results in serial plasma donations which had a median of 4-day intervals (and which were limited to twice weekly). In other words, the probability of obtaining a positive test result for a sample, along with a negative test result for the sample that was collected the following week from the same individual, was very low [13]. We parameterise the individual Weibull curves such that their delay from 5% to 95% takes 3.4 days. We set $k = 2$ in order to obtain a reasonable shape (see individual curves in Figure 2.1) and solve numerically to obtain $\gamma = 1.73$.

Delays from infectious exposure until 50% detectability and infectiousness are assumed to be normally distributed. The timescale (i.e. the maximum and minimum time values for which the area under the 'risk' curve is calculated) of calculations is made sufficiently large to ensure that the region of interest - where risk is nonzero - is fully accounted for.

The semi-mechanistic model is extended by allowing for inter-subject variability in viral doubling times and in (subject-level) mean per-virion infectiousness values. To this end, inter-subject variability is generated using truncated normal distributions. Truncation is necessary in order to impose reasonable bounds, e.g. doubling time must be positive. However, truncation shifts the distributional means and standard deviations from the parent (non-truncated) distributions. Since it is not possible to generate truncated normal distributions

[7] The general model and supporting code base support variability in all of the key model parameters, including the rate of transition from undetectable to highly-detectable. We focus on one type of variability here, since we believe this captures the key concepts at play and is more straightforward to analyse.

Stellenbosch University https://scholar.sun.ac.za

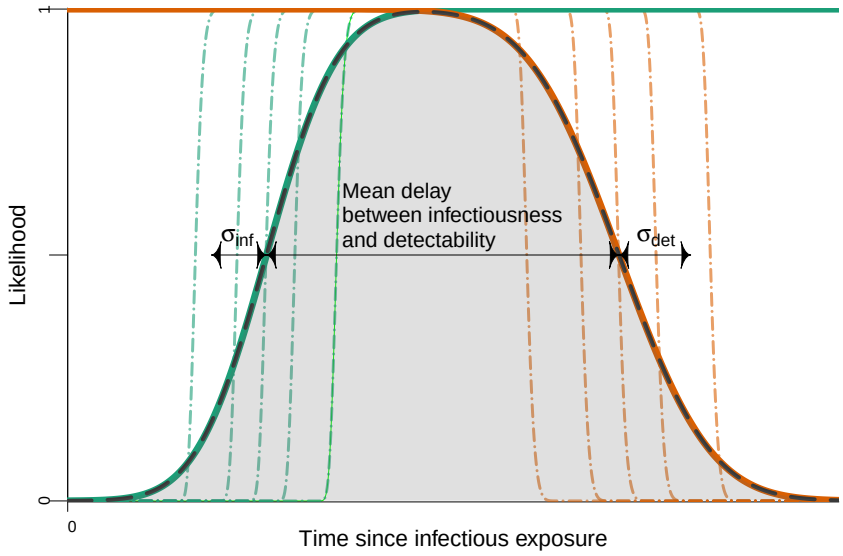


Figure 4.4: Illustration of the generalised approach to estimating the residual risk posed by an incident donor. Thin green curves show individual-level transitions from non-infectious to infectious, while the thick green curve shows the mean infectiousness at a given time following infectious exposure. Similarly, the thin orange curves show individual-level nondetection probabilities, while the thick orange curve shows the mean nondetection probability. σ_{inf} is the standard deviation of individual delays between exposure and becoming 50% infectious, while σ_{det} is the standard deviation of delays between exposure and becoming 50% detectable. The mean curves are generated from denser samples of underlying (i.e. not plotted) individual curves. The dashed grey line indicates the probability that the donated material will both test negative and cause an infection if transfused. The grey shaded area indicates the 'area' of risk (measured in days) posed, on average, by an incident donor.

Stellenbosch University https://scholar.sun.ac.za

with specified means and standard deviations through purely analytical methods, we use the R function `multisolve` (from the `rootsolve` package) to find pre-truncation parameters which lead to the desired post-truncation properties [98].

In both the generic model and the extended semi-mechanistic model, parameters are generated through inverse transform sampling of the applicable normal and truncated normal distributions. As well as being much more computationally efficient than pseudo-random sampling, and being easier to reproduce on other platforms, inverse transform sampling has the advantage of reducing the need to check, when using non-truncated normal distributions, for extreme outliers which could influence results.

Correlation in the general model is implemented according to the *order* in which subjects achieve 50% detectability and infectiousness. In other words, perfect positive correlation implies that subjects who become infectious the earliest (following exposure) also become detectable the earliest; if variability in infectiousness and detectability are identical, perfect positive correlation implies that every donor has the same delay between being 50% infectious and 50% detectable (differing only in their delays between infectious exposure and 50% infectiousness/detectability). Perfect negative correlation implies that the 'first' donors to become infectious (i.e. with the shortest delay between exposure and 50% infectiousness) are the last to become detectable (thus likely contributing a large amount to the overall risk) while the last donors to become infectious are the first to become detectable (and thus contribute a smaller amount to the overall risk than donors with early onset of infectiousness and late onset of detectability).

Different choices for how to parameterise correlation involve trade-offs in terms of presenting the results of the model. Parameterising correlation according to the magnitudes of delay times (or differences in delay times from the mean delay) creates a dependency between the levels of inter-subject variability and the overall strength of correlation. Normalising the expressions to remove dependency on other model parameters is equivalent to implementing correlation based on parameter order. Parameterising correlation according to the order of parameter values (e.g. delay to detectability) allows numerical access to the full dynamic range of the questions at hand, without depending on

Stellenbosch University https://scholar.sun.ac.za

detailed assumptions regarding the underlying parameters.

In the extensions of the semi-mechanistic model described here, correlation is implemented in terms of the order of the underlying parameters. Perfect positive correlation means that the subjects with the shortest viral doubling time λ have the smallest value of per-virion infectiousness. Perfect negative correlation means that subjects with the shortest doubling time have the highest value of per-virion infectiousness. Correlation can be 'dialed' continuously between 'perfect positive', 'no correlation' and 'perfect negative'.

Positive correlation is implemented by constructing a 'correlation matrix' in which each row/column position is assigned a value by calculating the probability density, in a normal distribution with zero mean, of the difference between the row and column number. The standard deviation of the normal distribution used determines the strength of correlation. Each matrix is normalised so that its elements sum to one. Matrices for generating negative correlation are constructed by reversing the column indices from the 'positive correlation' matrix. 'Correlated' risk estimates are made by calculating the 'individual-level' risk associated with every combination of parameter values (e.g. delays). The estimate for the total number of risk-day equivalents is then calculated as the weighted sum of 'individual-level' risk estimates. The weight of each estimate is given by the element in the corresponding correlation matrix with row number equal to the index of the first parameter and column number equal to the index of the second parameter.

4.3 Code base: a toolbox

Following infectious exposure, people may exhibit a variety of biomarker growth curves, undergo different types and numbers of tests at different times, and have varying levels of infectiousness over time. Code was developed for modelling the processes which underlie issues such as residual risk and infection dating. The code base allows models to be compared with the explicit knowledge of the assumptions encoded within them, and explore a variety of assumptions using one coherent and flexible tool.

The modelling framework is implemented in the R programming language

Stellenbosch University https://scholar.sun.ac.za

[99], using a number of publicly available packages. R was chosen for its range of statistical libraries, free availability, flexibility, and popularity among epidemiological and public health researchers (i.e. the intended user base for this project). Code was developed which may be used to explore the dynamics of systems which depend on test performance characteristics and biomarker dynamics, such as the residual risk of TTI's in screened blood products. While implementations chosen for the results presented in this work are necessarily limited by space and time, a number of extensions can be naturally incorporated into the analyses using the existing code. The code is also structured such that adding additional functionality would be relatively straightforward. The development of a flexible (may be re-purposed) and robust (avoids generating spurious results) code base required thoughtful design choices along with various checks and balances. The full code base can be found in a publicly available repository at: github.com/JemLukeBingham/biomarkers; snippets of code are shown in the rest of this chapter. See the `Readme.md` file in the repository for an overview of the repository structure.

The desired flexibility means users must be able to:

1. Specify the nature of the curves defining individual/mean biomarker levels, infectiousness, and/or detectability.

2. Specify the extent of variability in whichever parameters they choose, as well as the distributional assumptions for implementing variability.

3. Calculate the products of individual and/or mean-level curves.

4. Vary the strength of correlation between combinations of parameter values for parameters with variability (e.g. in different curves).

5. Calculate integrals of (areas under) the resultant product curves or related functions (e.g. calculating the expectation value of a posterior likelihood for a point-estimate of infection date).

Cumulative Weibull distributions were used to describe individual-level infectiousness and detectability. Cumulative Weibull distributions are usually parameterised from where they become non-zero; while this is maintained as an option, it is convenient to be able to specify where each curve reaches 50%,

Stellenbosch University https://scholar.sun.ac.za

since diagnostic delays and infectiousness estimates are often parameterised according to 50% detectability and 50% infectiousness. In order to generate cumulative Weibull distributions ('Weibull curves') with 50% values at specified locations, the position of each curve must be shifted by $\gamma \times \sqrt[k]{-\ln(0.5)}$. Functions are defined for generating a Weibull curve and shifting a curve such that the 50% value lies at a desired location; these are then wrapped these together in a convenience function which takes advantage of R's natural vectorisation:

```
shift_to_half_likelihood_weibul <- function(scale,
                                            shape,
                                            maxSensitivity = 1){
  return(scale*(-1*log(1/(2*maxSensitivity)))^(1/shape))}


sensitivity_weibul <- function(x,
                               scale,
                               shape,
                               position,
                               falsePositiveRate = 0,
                               maxSensitivity = 1){
  return(ifelse(x-position<0,
                falsePositiveRate,
                maxSensitivity*(1-exp(-((x-position)/scale)^shape))))}


individual_sensitivity_weibul <- function(times,
                                          scale,
                                          shape,
                                          delay){
  return(sensitivity_weibul(x=times,
                            scale = scale,
                            shape = shape,
                            position =
                            delay -
```

Stellenbosch University https://scholar.sun.ac.za


```
                                    shift_to_half_likelihood_weibul(shape=shape,
                                                                    scale=scale)))}
```

The function ('individual_sensitivity_weibul') can then be called with vectors of scale and/or shape and/or delay parameters. Provided the vectors are the same length, the function will (with R's natural vectorisation) produce curves over the 'times' vector for each of the parameters in the vectors. This method is faster and more elegant than simply writing a 'for' loop in R to calculate all of the individual-level curves. It also allows the user to specify whichever parameter combinations they choose, without being tied to particular methods for simulating variability.

Using the existing code, variability may be simulated using either inverse transform sampling or pseudo-random sampling (with R's built-in random number generators), from normal distributions or truncated normal distributions. Truncation avoids spurious outputs resulting from unrealistic individual-level parameter values (such as 'individual-level' curves with substantial probabilities of testing positive at the time of infectious exposure, or an 'individual' with a negative per-virion infectiousness value). It also makes results more robust to arbitrary choices such as the number of individual-level curves to simulate when using inverse transform sampling (which causes the range of parameter values to increase when the number of 'individual-level' curves is increased).

```
get_pos_cum_norm = function(n,
                            mean_center_position,
                            sd_size){
  myseq <- seq(1/(2*n),1-1/(2*n),1/n)
  positions<-qnorm(myseq,
                   mean=mean_center_position,
                   sd=sd_size )
  return(positions)}


get_pos_cum_norm_trunc = function(n,
                                  mean_center_position,
```

Stellenbosch University https://scholar.sun.ac.za


```
                                    sd_size,
                                    minVal = -Inf,
                                    maxVal=Inf){
  myseq <- seq(1/(2*n),1-1/(2*n),1/n)
  positions<-qtruncnorm(myseq,
                        mean=mean_center_position,
                        sd=sd_size,
                        a = minVal,
                        b = maxVal)
  return(positions)}
```

In order to be able to specify the realised mean of a truncated normal distribution with non-symmetric truncation (relative to the pre-truncation mean), code was written utilising R's equation solver `multiroot`:

```
diff_realvsspec_mean_trunc_norm <- function(x,
                                            parms=c(desired_mean,
                                                    input_sd,
                                                    minVal,
                                                    maxVal)){
  desired_mean= parms[1]
  input_sd=parms[2]
  minVal=parms[3]
  maxVal=parms[4]
  return(abs(parms[1] - etruncnorm(a = parms[3],
                                   b=parms[4],
                                   mean=x,
                                   sd = parms[2])))}


get_pos_cum_norm_trunc_specmeans <- function(n,
                                             desired_mean,
                                             input_sd,
                                             minVal = -Inf,
```

Stellenbosch University https://scholar.sun.ac.za


```
                                                maxVal=Inf,
                                                relativeTolerance=1e-6,
                                                iterationLimit=10000,
                                                printPreTruncMean = FALSE){
  if(!(desired_mean<maxVal & desired_mean>minVal)){
    stop("You won't be able to find a truncated
         normal with the combination of mean and max/min
         values which you've specified.")}
  if(input_sd == 0){if(printPreTruncMean){print(desired_mean)}
    return(rep(desired_mean,n))}
  preTruncMean = multiroot(diff_realvsspec_mean_trunc_norm,
                           parms = c(desired_mean = desired_mean,
                                     input_sd = input_sd,
                                     minVal = minVal,
                                     maxVal = maxVal),
                           start=desired_mean,
                           maxiter = iterationLimit,
                           rtol= relativeTolerance)['root']
  if(printPreTruncMean){print(preTruncMean)}
  return(get_pos_cum_norm_trunc(n = n,
                                mean_center_position = preTruncMean,
                                sd_size = input_sd,
                                minVal = minVal,
                                maxVal = maxVal))}
```

Calculating the product of two curves at each time point (representing joint likelihoods) is fairly straightforward in R; we include a wrapper function `generateProductCurve` to make the code more readable.

Correlation is implemented by calculating the risk resulting from each pair of parameters (or parameter combinations), then weighting the sum of these individual-level risks according to the entries in a 'correlation matrix'. The correlation matrix may be specified arbitrarily, so long as it is normalised (its entries sum to 1) and has the same number of rows and columns as the number of values for each variable parameter. Code is currently available for generating

Stellenbosch University https://scholar.sun.ac.za

correlation matrices according to the order of parameter values, or according to differences between individual parameter values and the mean parameter values. In the simpler case (where correlation is generated according to order) the user must supply the number of values for each parameter and a measure of the strength of correlation. The function which generates the correlation matrices checks some basic features of the input, then outputs a normalised matrix of weights. Negative correlation is achieved by reversing the order of the rows or columns in the matrices.

```
get_Cor_Matrix_TimeOfEvent <- function(n,
                                       sdCorrelation,
                                       eventTimes1,
                                       sd1 = sd(eventTimes1),
                                       eventTimes2,
                                       sd2 =sd(eventTimes2),
                                       illustrate = FALSE){
  mean1 <- mean(eventTimes1)
  mean2 <- mean(eventTimes2)
  correlationMatrix <- matrix(nrow=n, ncol=n)
  normalise <- 0
  if (sdCorrelation ==0){
    if(length(eventTimes1)!= length(eventTimes2)){
      print("Whoah! This thing wasn't built to handle perfect
            correlation of different-size event-time vectors.
            What would that even mean?")
    }
    return(diag(length(eventTimes1))/length(eventTimes1))
  }
  correlationMatrix <- matrix(nrow=length(eventTimes1),ncol=length(eventTimes2))
  normalise <- 0
  for (i in 1:nrow(correlationMatrix)){
    for (j in 1:ncol(correlationMatrix)){
      correlationMatrix[i,j] <- dnorm(abs((eventTimes1[i]-mean1) /
                                          sd1 - (eventTimes2[j]-mean2)/sd2),
```

Stellenbosch University https://scholar.sun.ac.za


```
                                        sd=sdCorrelation,
                                        mean=0)
    }
  }

  correlationMatrix <- correlationMatrix / sum(correlationMatrix)

  if(illustrate){
    plot(eventTimes1,
         eventTimes2,
         type='n')
    for (i in 1:nrow(correlationMatrix)){
      for (j in 1:ncol(correlationMatrix)){
        points(eventTimes1[i],
               eventTimes2[j],
               cex=correlationMatrix[i,j]*(5/max(correlationMatrix))*2/sqrt(n))
        textxy(eventTimes1[i],
               eventTimes2[j],
               paste("(",i,",",j,")"))
      }
    }
    persp(eventTimes1,
          eventTimes2,
          correlationMatrix,
          theta = 110,
          phi = 10,
          ticktype = "detailed",
          expand = 0.5,
          shade = .2)

    print(max(abs(correlationMatrix - t(correlationMatrix)) /
                (correlationMatrix/2+t(correlationMatrix)/2)))
    }
  return(correlationMatrix)}
```

Stellenbosch University https://scholar.sun.ac.za


```
get_Cor_Matrix_Order <- function(n,
                                 sdCorrelation,
                                 illustrate = FALSE){
  if (sdCorrelation ==0){
    return(diag(n)/n)
  }
  correlationMatrix <- matrix(nrow=n,
                              ncol=n)
  normalise <- 0
  for (i in 1:nrow(correlationMatrix)){
    for (j in 1:ncol(correlationMatrix)){
      correlationMatrix[i,j] <- dnorm(abs((i-j)),
                                      sd=sdCorrelation,mean=0)
    }
  }
  correlationMatrix <- correlationMatrix / sum(correlationMatrix)
  if(illustrate){
    persp(1:n,
          1:n,correlationMatrix,
          theta = 110,
          phi = 20,
          ticktype = "detailed",
          expand = 0.5,
          shade = .2,
          xlab = "Index of parameter 1",
          ylab = "Index of parameter 2",
          zlab = "")
    print(max(abs(correlationMatrix - t(correlationMatrix)) /
                (correlationMatrix/2+t(correlationMatrix)/2)))
    }
  return(correlationMatrix)}
```

Stellenbosch University https://scholar.sun.ac.za

Single-variable definite integrals in R are calculated using the AUC (area under the curve) function from the DescTools package [100]:

```
naive_risk = AUC(times,naive_likelihood).
```

Functions were created to calculate and combine the constituent components of the semi-mechanistic model.

```
generateViralGrowthCurve<- function(C0=1,
                                    dblT,
                                    eclipse=0,
                                    times){
    return(C0*2^((times-eclipse)/dblT))}


calc_p_nondetect <- function(VL,
                             X50,
                             X95,
                             Z,
                             RNAPerVirion = 2){
  return(1 - pnorm(Z*(log(VL*RNAPerVirion/X50))/(log(X95/X50)),
                   mean=0,
                   sd=1))}


calculate_P_infect <- function(VL,
                               vol,
                               pV){
  return( 1 - exp(-VL * vol * pV) )}


calc_Weusten_risk <- function(dblT,
                              pV,
                              X50 = 2.7,
                              X95 = 18.4,
```

Stellenbosch University https://scholar.sun.ac.za


```
                                C0=1/1000,
                                eclipse = 10,
                                maxTime = 30,
                                detail = 4,
                                vol = 20,
                                poolSize = 1,
                                volume_correction = FALSE,
                                illustrate = FALSE){
  # defaults to the values used by Weusten et al (2011),
  # i.e. the procleix Ultrio essay HIV-1; IDNAT
  times = seq(0,maxTime,1/detail)
  viralLoads = generateViralGrowthCurve(C0 = C0,
                                        dblT = dblT,
                                        eclipse = eclipse,
                                        times = times)
  viralLoads2 = viralLoads / poolSize
  if(volume_correction==TRUE){viralLoads2 = viralLoads2 / 2}
  nondetectionProbability = calc_p_nondetect(VL = viralLoads2,
                                             X50 = X50,
                                             X95 = X95,
                                             Z = qnorm (0.95))
  infectionProbability = calculate_P_infect(VL = viralLoads,
                                            vol = vol,
                                            pV = pV)
  if (nondetectionProbability[length(times)]>.001){
    print("nondetection probability too
          large - need more time")
  }
  risk = nondetectionProbability*infectionProbability
  if (illustrate){
    plot(times,
         nondetectionProbability,
         xlab = "Time (days)",
         ylab= "Likelihood",type='l')
```

Stellenbosch University https://scholar.sun.ac.za


```
    lines(times, infectionProbability)
    lines(times, risk)
    points(times, risk)
  }
  return(AUC(times,risk))}
```

Risk estimates from the base semi-mechanistic model may be calculated using a single line of code.

```
calc_Weusten_risk(dblT = 0.85, pV = 0.217, poolSize = 2)
```

In addition, users must be confident that the outputs of the simulations are caused by the assumptions they encoded, rather than by anomalies in the underlying code. For example, in our simulations infectiousness and detectability levels at the earliest/latest time points must not be significantly different from to their maximal or minimal values, to ensure that the risk is not underestimated as a result of simply not including some of the risk-time in the calculation. Users must also be confident that the results of simulations are reproducible. In order to address these challenges we include various checks within the key functions, as well as snippets of commented-out code for testing the behaviour of individual functions. Some functions, including the functions for creating correlation matrices, and the function for calculating risk using the base semi-mechanistic model (i.e. as described by Weusten et al. [14]), include an option to 'illustrate' the results, causing the output of the calculations to be plotted:

```
calc_Weusten_Risk(dblT = 0.85,
                  pV = 0.217,
                  poolSize = 2,
                  illustrate = TRUE),
```

or

```
n = 20
get_Cor_Matrix_Order(n = n,
                     sdCorrelation = n/2,
                     illustrate = TRUE).
```

Stellenbosch University https://scholar.sun.ac.za

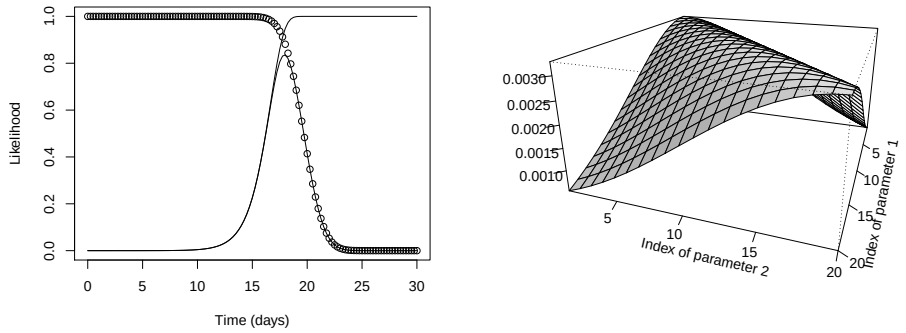


Figure 4.5: Left panel: Optional illustration produced by the `calc_Weusten_Risk` function, showing the infectiousness and detectability of a blood product (one unit of RBC component) from a hypothetical donor, as functions of time since the onset of exponential viral growth. Right panel: Optional illustration produced by the `get_Cor_Matrix_Order` function. z-axis values represent the (normalised) weights applied to each pair of parameter values, when weighted sums of individual contributions to the average risk are calculated. x-axis and y-axis values represent indices which specify positions in two vectors of parameter values (e.g. a vector of per-virion infectiousness values and and a vector of viral doubling times). The relative probability of a recently-infected hypothetical donor exhibiting a combination of, for example, the third value in the vector of per-virion infectiousness values, and the seventh value in the vector of viral doubling times, will be given by the z-axis value at $x = 3, y = 7$.

Code structure

The code base is structured as follows: all functions as well as default parameter values (including estimated parameters such as mean viral doubling time as well as parameters such as plot colour schemes) are defined in separate files ('functions.R' and 'parameters.R'). Scripts for generating and displaying specific results are created independently and begin by running the same function and parameter scripts. Following many function definitions are brief tests for the functions; currently these must be manually checked by un-commenting the relevant lines and checking the output. Next to parameter definitions are (where applicable) digital object identifiers (DOIs), paper titles, and/or author names for the sources of the parameter estimates.

We chose to make names informative and naturally readable, rather than brief. Functions are named using verbs while outputs and parameters are named us-

Stellenbosch University https://scholar.sun.ac.za

ing nouns. The use of camelCase versus separating 'words' in names using underscores is still inconsistent, though newer code uses underscores.

Results are generated independently for each model. Each set of results and accompanying figures is generated in a single script. The overall workflow is:

1. Run scripts to define functions and default parameters.
2. Define the parameters for implementing variability (e.g. number of curves to use in each simulation, length of simulations, etc), and overwrite default parameters for individual curves if desired.
3. Create matrices with which to generate correlation between chosen parameters.
4. Create distributed parameter values. This defines the size of the results; more distributed values means more potential combinations.
5. For each combination of parameter values and correlation levels calculate risk.

The code base described here constituted a use-able tool which can be adapted with relative ease to address a wide range of questions related to diagnostic test performance.

Stellenbosch University https://scholar.sun.ac.za

Chapter 5

Results: calculating the effective size of the window period

Comparing results between models

The amount of TT-HIV risk posed by a hypothetical incident donor - analogous to the window period, occurring in units of days, and referred to as 'risk days per incident donor' or simply 'risk per incident donor' - is calculated using both the generic model and the extended semi-mechanistic model. Results are compared to equivalent estimates using methods from the IWP model (i.e. the mean delay between infectiousness and detectability) and the original semi-mechanistic model [20, 14]. The number of 'risk-day equivalents' posed by an incident donor is directly proportional (once donor incidence is included) to the number of resultant infections among recipients, subject to the assumption that donation rates are unaffected by infectious exposure[1]. Results are reported using the summary metric of 'risk-day equivalents', which captures the interaction of the biology of transmission with the biology and technology of *detectable* infection.

[1]To be precise, the total risk is proportional to the risk per incident donor in risk-day equivalents when the incidence rates are uniform and equal during the period over which incidence is measured, and the period during which the risk is substantially different from zero. In the new IWP for example, one can relax the assumption of fully uniform donation rates, and assume only that donation rates are uniform between the time of infectious exposure and antibody conversion.

Stellenbosch University https://scholar.sun.ac.za

5.1 Risk estimates using the generic model

The number of risk-day equivalents per incident donor calculated using the generic model is always larger than (or equal to) the mean delay between 50% infectiousness and 50% detectability (also referred to simply as the 'mean delay') (Figures 5.1 - 5.3).

The relative difference between the generic model and the mean delay is larger when the mean delay is 'near' to zero, while the generic model predicts risk close to the mean delay when the magnitude of the mean delay is large. The *scale* at which the risk under the generic model differs substantially from the mean delay depends on the magnitude of inter-subject variability. The bottom panel of Figure 5.1 shows results from the generic model with relatively large inter-subject variability while the top panel shows results with lower levels of inter-subject variability.

Correlation affects risk estimates substantially. Perfect positive correlation results in risk estimates near to the mean delay, differing meaningfully (in relative terms) only when the mean delay becomes small or negative. Perfect negative correlation generates larger risk estimates than perfect positive correlation for equivalent levels of inter-subject variability and mean delays. Changing the level of correlation approximately continuously from positive to negative causes risk estimates to increase approximately continuously (not plotted), and estimates assuming imperfect correlation lie between estimates made using perfect positive and estimates made using perfect negative correlation.

Each risk estimate with inter-subject variability is a weighted sum of risks represented by pairs of 'individual-level' curves. Each pair of individual-level curves, made up of one 'infectiousness' curve and one 'non-detection probability' curve, has an associated 'delay' (i.e. between the times at which the 'individual-level' infectiousness and detectability curves reach 50%).

Correlation in this model means assigning probabilities to pairs of individual-level curves. With no correlation, all weights are equal - i.e. every possible pair of individual-level infectiousness and nondetectability curves is equally likely.

In the generic model with perfect positive correlation and equal levels of inter-subject variability (in the timing of infectiousness/detectability), the contribution to the total risk is the same for every 'individual-level' pair of curves. Be-

Stellenbosch University https://scholar.sun.ac.za

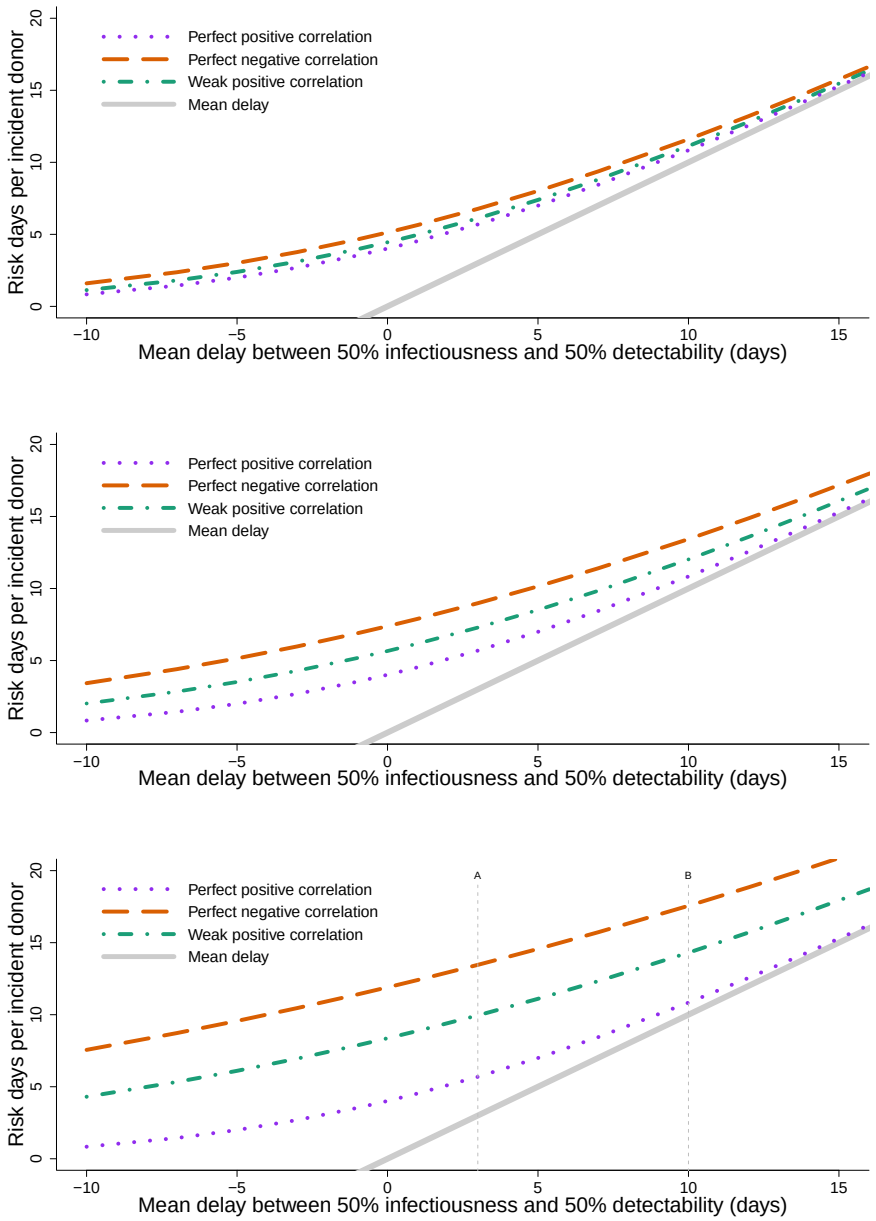


Figure 5.1: Risk per incident donor (measured in risk-day equivalents) in the implemented version of the general model, with relatively high inter-subject variability in the timing of infectiousness and detectability ($\sigma_1 = \sigma_2 = 13$ days. Risk estimates depend strongly on the assumed degree of correlation between the times at which subjects become infectious and detectable. The largest risk estimates arise with perfect negative correlation - where subjects who become infectious first become detectable last and vice versa - which maximises heterogeneity in the contribution of individual donors to overall risk, while perfect positive correlation generates risk estimates which differ from the mean delay only when the mean delay approaches zero. The effect of variability become smaller when the mean delay becomes large.

Stellenbosch University https://scholar.sun.ac.za

cause every 'individual-level' pair of curves is contributing the same amount to risk, the risk estimate is determined purely by the shapes of the pairs of individual-level curves (and the individual-level delays between infectiousness and detectability, which are all equal to the mean delay). Since the cumulative Weibull curves used are nearly symmetrical around their 50% values [2], when the mean delay is long relative to the magnitude of intra-subject variability (i.e. the duration of an individual curve's transition from 0 to 1 or vice-versa) the contribution to risk from a particular pair of curves is very close to the delay between their 50% infectiousness and 50% detectability values.

The difference between the models is larger when the mean delay between infectiousness and detectability is small, and when the correlation is weak or negative. Results assuming perfect positive correlation between the times at which subjects become infectious/detectable deviate from the mean delay only when the mean delay is small (relative the the magnitude of inter-subject variability) or negative (in which case the actual amount of residual risk depends in detail on the shapes of individual detectability and infectiousness curves). On the other hand, when correlation is weak or negative the difference between the models is larger when there is more inter-subject variability.

Furthering this approach, we calculate the overall risk using different levels of inter-subject variability and mean delays, and compare the total magnitude of the variability $\left(\sqrt{\sigma^2_{\text{infect}} + \sigma^2_{\text{detect}}}\right)$ with the resulting estimates (Figures 5.2 and 5.3). Interesting to note is that estimates under perfect positive correlation decrease with increasing inter-subject variability, up until the point where $\sigma_1 = \sigma_2$. This occurs because both delays (from infectious exposure to infectiousness and detectability respectively) are drawn from normal distributions, so that with *perfect* correlation, inter-subject variability in delays from infectiousness *to* detectability is minimised when both delay distributions have the same standard deviation (which is equivalent to the model without variability, since variability only in eclipse periods does not affect the results). However, as discussed

[2] The 'true' curves for a particular individual might not be symmetric - while this could be easily incorporated into the model, the details of this non-symmetry are in general impossible to parameterise, would have a substantial effect on model results (meaning that results would be 'baked-in' to specific choices regarding individual infectiousness and detectability profiles, rather than reflecting the general properties of intra- and inter- subject variability), and would obscure comparisons with the 'base' model (i.e. the 'window period', as in Figure 3.1).

Stellenbosch University https://scholar.sun.ac.za

in section 4.1, *perfect* positive correlation is highly unlikely to occur in reality, and a modest departure from perfect positive correlation (black points in Figures 5.2 and 5.3) leads to qualitative agreement with Figure 5.1, where negative correlation and increased variability imply higher levels of risk.

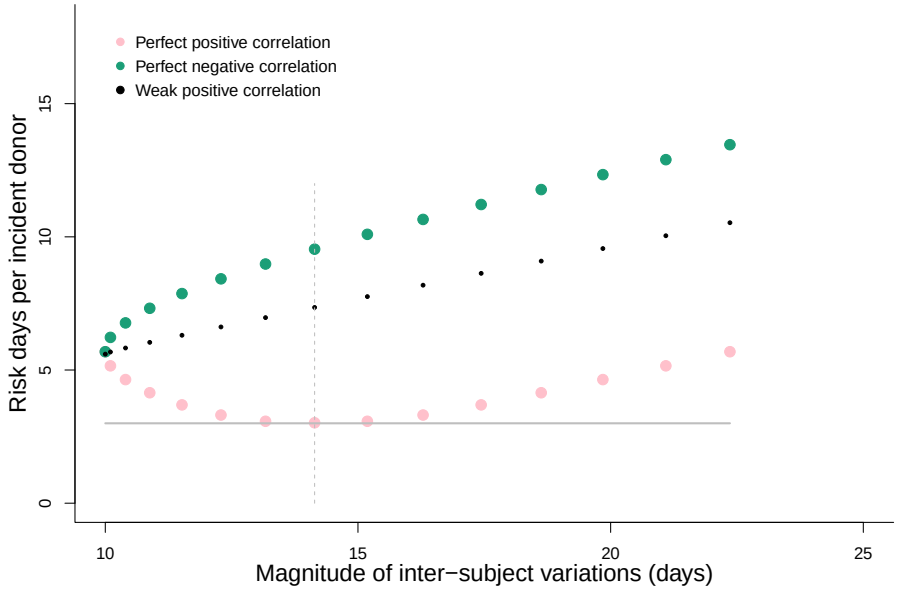


Figure 5.2: Risk per incident donor using the generic model, for different 'total magnitudes' $\left(\sqrt{\sigma_1^2 + \sigma_2^2}\right)$ of inter-subject variability in delays to 50% infectiousness and detectability. σ_1 is kept fixed at 10 days, while σ_2 is varied between 0 (on left side of the plot) and 20 (at the right-most points). The mean delay is kept fixed at 3 days. Calculations with perfect positive correlation are in general lower than those with perfect negative correlation.

Stellenbosch University https://scholar.sun.ac.za

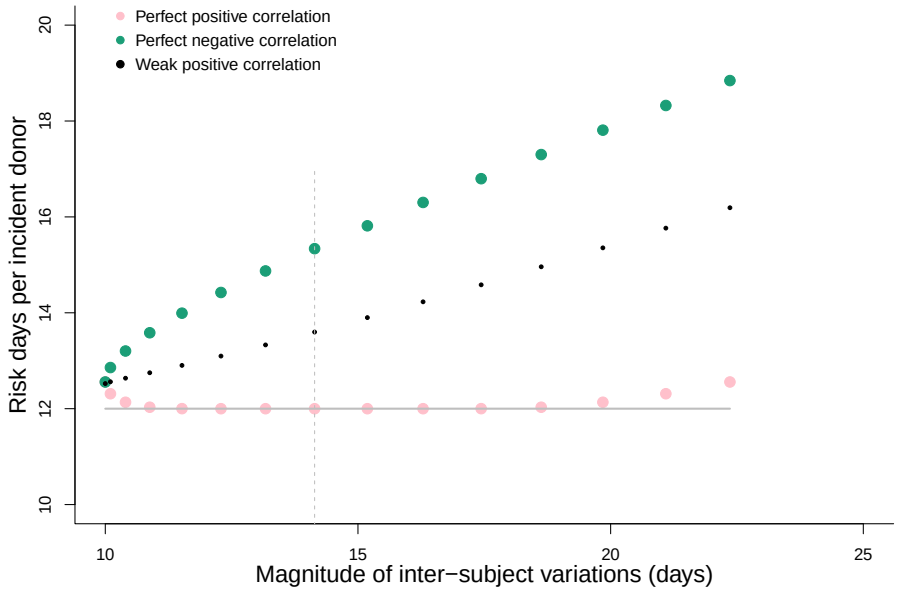


Figure 5.3: Risk per incident donor using the generic model, for different 'total magnitudes' $\left(\sqrt{\sigma_1^2 + \sigma_2^2}\right)$ of inter-subject variability. σ_1 is kept fixed at 10 days, while σ_2 is varied between 0 (on left side of the plot) and 20 (at the rightmost points). The mean delay is kept fixed at 12 days. Calculations with perfect positive correlation are in general lower than those with perfect negative correlation.

5.2 Risk estimates using the extended semi-mechanistic model

Extending the semi-mechanistic model to incorporate inter-subject variability in viral doubling time and per-virion infectiousness, the risk per incident donor (in risk-day equivalents) is calculated and compared to results from the base model. Calculations using the base model are made using the realised mean parameter values from the extended model.

In all cases, we generate 'individual' per-virion infectiousness values (pV) and viral doubling times (λ) from truncated normal distributions such that $0.0345 \leq \text{pV} \leq 0.9000$ and $0.224 \leq \lambda \leq 2.275$ days. The peaks of the distributions are set to the mean values used by Weusten et al. ($\bar{\text{pV}} = 0.217$ and $\bar{\lambda} = 0.85$) [14]. Specifying the bounds of the distributions means that the realised means and standard deviations cannot in general be specified exactly. At

Stellenbosch University https://scholar.sun.ac.za

the limits of the achievable values, the distribution becomes flatter, and the pre-truncation mean may shift out of the truncation range. In our case, the peak of the distribution sits below the mean of the bounds of the distribution, and increasing variability in either parameter results in an increased mean value. Due to the linear relationship between doubling time and risk-days in the underlying model, the effect of doubling-time variability may be isolated if we divide risk estimates using the nuanced model (which are essentially weighted means of the risks posed by each characteristic 'subject'-level curve) by the realised mean doubling time.

We begin by examining the effects of variability in doubling time and per-virion infectiousness on the total risk-days per incident donor, for blood screened by an algorithm consisting of a single Ultrio Plus test in ID-NAT configuration (Figures 5.4 and 5.5).

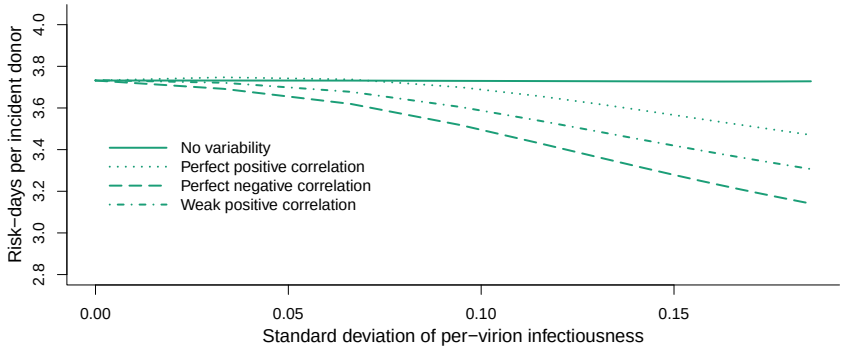


Figure 5.4: The effect of variability in per-virion infectiousness p_V on risk per incident donor, for RBC screened for HIV using a single Ultrio Plus assay with a pool size of 2. Perfect positive correlation (i.e. subjects with the lowest p_V have the shortest doubling time λ, dotted line) generates risk estimates slightly larger than the base model, while perfect negative correlation (dashed line) generates smaller risk estimates. As expected, both these effects are stronger with more variability. The mean p_V value is kept constant at 0.217 [14].

Stellenbosch University https://scholar.sun.ac.za

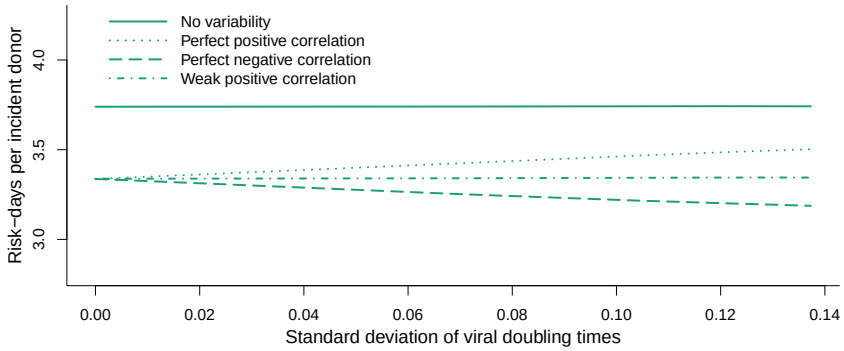


Figure 5.5: The effect of variability in viral doubling times λ on risk per incident donor, for RBC screened for HIV using a single Ultrio Plus assay with a pool size of 2. Perfect positive correlation (i.e. subjects with the lowest p_V have the shortest doubling time λ, dotted line) generates risk estimates slightly larger than the base model, while perfect negative correlation (dashed line) generates smaller risk estimates. As expected, both these effects are stronger with more variability. The mean p_V value is kept constant (within 0.1%) at 0.217 [14, 89]

Note that in Figure 5.5 the total risk estimates with/without variability do not match up when $\sigma_\lambda = 0$. This is because $\sigma_{p_V} \neq 0$ and the non-linear dependence of total risk on p_V means that the model estimate using the mean value is larger that the mean of the 'individual' model estimates.

In line with our analysis of the general model, we wish to explore the effects of reduced sensitivity - in this case analytic sensitivity, rather than explicitly lengthened diagnostic delay - in the context of inter-subject variability.

Stellenbosch University https://scholar.sun.ac.za

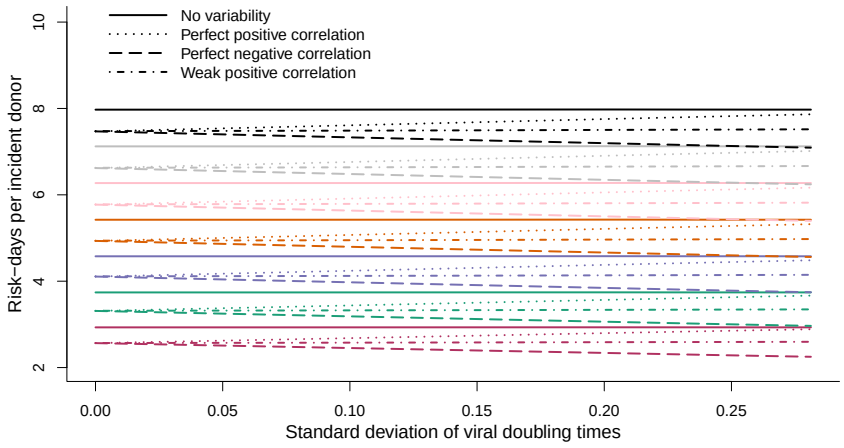


Figure 5.6: The effects of variability in viral doubling time, for different levels of test sensitivity (or, equivalently, minipool size), under the extended semi-mechanistic model. Each colour represents a factor-reduction in test sensitivity (or increase in pool size) relative to a Procleix Ultrio assay testing for HIV in ID-NAT configuration: 1 (maroon lines), 2 (green lines), 4 (purple lines), 8 (orange lines), 16 (pink lines), 64 (grey lines), and 96 (black lines). 'No variability' (solid lines) refers to the model without any variability at all, while the other lines are generated with a fixed level of variability in per-virion infectiousness (sd = 0.75) [14, 83].

Stellenbosch University https://scholar.sun.ac.za

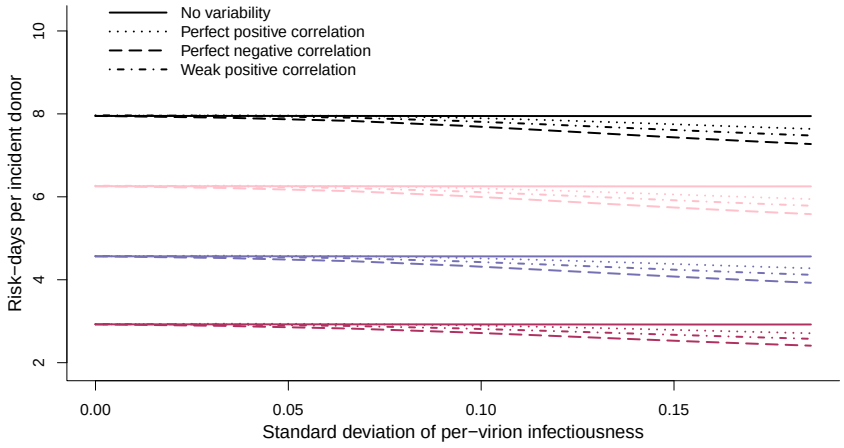


Figure 5.7: The effects of inter-subject variability in mean 'per-virion infectiousness', for different levels of test sensitivity (or, equivalently, minipool size) under the extended semi-mechanistic model. Each colour represents a factor-reduction in test sensitivity (or increase in pool size) relative to a Procleix Ultrio assay testing for HIV in ID-NAT configuration: 1 (maroon lines), 4 (purple lines), 16 (pink lines), and 96 (black lines). 'No variability' (solid lines) refers to the model without any variability at all, while the other lines are generated with a fixed level of variability in viral doubling time (sd $= 1$).

Stellenbosch University https://scholar.sun.ac.za

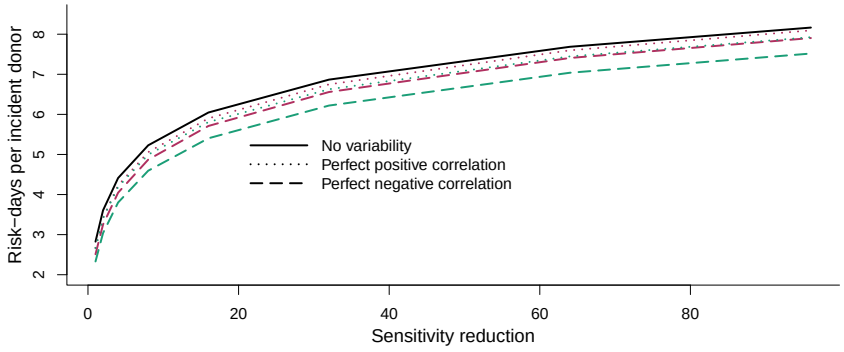


Figure 5.8: The effects of inter-subject variability in viral doubling times, for different levels of test sensitivity/pool size, under the extended semi-mechanistic model. All estimates assume a mean viral doubling time of 0.85 days, and a mean 'per-virion infectiousness' of 0.217. Green lines show risk estimates assuming a standard deviation in viral doubling times of 1 day, while maroon lines show risk estimates assuming a standard deviation in doubling time of 0.14 days. 'No variability' (solid lines) refers to the model without any variability at all, while the other lines are generated with a fixed level of variability in 'per-virion infeciousness' (sd $= 0.75$).

Stellenbosch University https://scholar.sun.ac.za

Chapter 6

Discussion and conclusion

6.1 Discussion

Modelling the performance of tests for disease and infection has proved valuable in a number of fields, including blood safety and infection dating. In this work, we have taken steps to develop a largely unexplored aspect of test performance modelling: inter-subject variability coupled with time-varying test performance. We have described a broad framework through which these factors may be considered, and subsequently applied this framework to relevant questions in infection dating (see Grebe et al. [16] for details) and estimating the residual risk of transfusion-transmitted HIV infections.

What are the implications of adding this particular type of nuance to residual risk estimation models? The main take-away from this work is, unsurprisingly, that *context* determines how much nuance is appropriate to include in a model. Nonetheless, the discovery of contexts where incorporating variability both affects net risk estimates, and extends model usefulness to new domains, demonstrates our primary message: whenever models are constructed which include or rely on descriptions of test performance characteristics, researchers should begin with generalised approaches. Only after the implications of nuanced descriptions have been explored may responsible decisions regarding model structure be made. The generalised approach taken in this work provides a way of confirming whether/when a more nuanced model is called for, and when one can confidently use a simpler model: the more general analysis

Stellenbosch University https://scholar.sun.ac.za

leads one to a clearer understanding of the domains of validity of the simpler analyses.

Context affects the results in different dimensions; firstly, the underlying model structure naturally has a significant effect on risk estimates, and on the effects of adding variability to the models. Lengthening the delay between infectiousness and detectability in the generic model reduces the impact of the nuanced approach. Although this relationship is less straighforward in our extension of the semi-mechanistic model[1], a similar trend appears in the relative risks of the model with/without inter-subject variability appears, whereby larger differences between the more-simplistic base model and our extended model coincide with longer mean delays between the infectiousness and detectability.

Results from the generic model suggest that in domains with large delays between the onset of infectiousness and the onset of detectability, such as TT-HIV risk estimation with antibody-only test algorithms, simplistic model assumptions prove well-justified, and results are not strongly affected by the addition of nuance. When the delay between the onset of infectiousness and the onset of detectability is small or negative, nuanced approaches lead to substantial differences in results.

The dynamics of the semi-mechanistic model and accompanying extensions should be viewed in the context of the underpinning viral growth model. For example, the linear dependence of risk on the viral doubling time occurs because every time-dependent input to the model is in the form $2^{\frac{t}{\lambda}}$. As such, changing λ amounts to re-scaling the time axis. Since the total risk-days estimate is an integral over all possible times, linearly re-scaling the time axis re-scales the total risk-days estimate. Extrapolating to zero doubling time yields the estimate of approximately zero risk, as expected (since zero doubling time means the viral load would jump immediately from negligibly infectious to reliably detectable). Reducing the sensitivity of the screening algorithm (e.g. by increasing the pool size) by factors of two corresponds to approximately arithmetic increases in total risk estimates.

Inter-subject variation in per-virion infectiousness p_V causes reduced risk estimates due to its non-linear relationship with individual-level risk. Since the

[1]For example, the level of inter-subject variability in per-virion infectiousness is tied to the mean delay.

Stellenbosch University https://scholar.sun.ac.za

total risk per seroconverting donor is essentially a weighted sum of 'individual' risks, the individual risks have the same dependency on p_V as the total risk in the base model. Changes in p_V have a larger effect on total risk for small p_V values, so that the mean of two risk estimates with different p_V values will be skewed towards the risk estimate under the smaller value.

Positive correlation in the extended semi-mechanistic model means that small viral doubling times imply small p_V values - the safest possible combination, since low doubling time causes a rapid transition from infectiousness to detectability, while low p_V means detectability is reached at lower infectiousness levels. On the other hand, negative correlation implies that individuals with large doubling times, who have a slow transitions between infectiousness and detectability (and hence wider areas of risk), also become detectable at higher levels of infectiousness. The unequal distribution of individual-level contributions to risk causes an increase in the total risk per seroconverting donor. Negative correlation, where the effects of doubling time and p_V on risk counteract one another at the individual level, results in the lowest risk estimates.

Whether positive or negative correlation is more realistic will depend on the details of the pathogen, tests, and epidemiology (e.g. average time-since infection) in question and is likely in many cases to remain a mystery. Early dynamics of HIV infections have proven challenging to characterise [31, 43, 40]. The available evidence suggests that in established HIV infections, less infectious viral strains are more abundant than the initial (more-infectious) founder strain. On the other hand, it is believed that the early stages of adaptive immune response target primarily non-founder strains of the virus [40].

Clearly the generalised framework may include models whose complexity does *not* add value to end results, and in these circumstances simpler models should be used. Indeed, the analyses presented in this work provide a way of specifically confirming that simpler models are adequate, when applied in appropriate contexts. However, as we have shown, there are circumstances where a more nuanced approach *does* have an impact on results. Work which relies on modelling the performance of diagnostic test characteristics should therefore be conceptualised within the generalised framework; models may then be refined according to which aspects of the generalised viewpoint are deemed useful. This represents a non-trivial departure from the status-quo for the epidemio-

Stellenbosch University https://scholar.sun.ac.za

logical modelling of test properties.

6.2 Directions for future work

A number of avenues exist through which the modelling approaches described in this work might be extended. These include:

- exploring the implications of non-uniform timing of blood and/or tissue donations relative to the timing of donor infectious exposure.
- modelling the residual risk of other transfusion-transmitted infections (e.g. HCV, HBV, West Nile Virus, etc.), and extending incidence-estimation approaches such as those described by Grebe et al [16] to incorporate other pathogens of interest.
- extending established residual risk estimation approaches to contexts in which blood banking practices which fall short of WHO best-practice guidelines.
- modelling the implications of a pathogen-inactivation technology for residual risk of transfusion-transmitted infections.

Donation timing

If one wanted to explore the implications of non-homogeneous donation timing on residual risk, this would fit neatly into the general modelling framework we have proposed, and would require only minor modifications to the existing model and code. In essence, one could apply a time-dependent weight to the individual-level risk curves capturing an assumption regarding the effect of infectious exposure on donation rates. This would require few additional lines of code to implement. More detailed assumptions would also be possible - for example, the modification to donation rates following infectious exposure could be correlated with other model parameters.

When making incidence estimates using passive surveillance systems, where infected individuals must take action in order to be tested (whether directly

Stellenbosch University https://scholar.sun.ac.za

seeking a test, or indirectly via seeking healthcare), non-uniformity in the probability of sample collection among infected individuals, relative to the timing of infectious exposure, may have substantial impacts on results. The time-varying probability of an infected individual having their sample collected may also depend on non-homogeneous (in time, space, etc.) factors such as access to healthcare and social perceptions regarding the disease in question. Similarly to the issue of non-uniform donation timing, such scenarios could be neatly incorporated into the modelling framework and code base described here.

Other TTI's

Qualitatively realistic modelling of transfusion transmission risk from other pathogens should follow a similar pattern to what has been described in this work. The accuracy and precision with which models can be parameterised varies drastically between diseases, depending generally on incidence, disease life-cycle, and the extent of efforts to develop and parameterise tests. For diseases such as HIV, where the details of the transition from non-detectable to detectable (and non-infectious to infectious) are difficult to specify, despite substantial efforts in this direction, one could modify the shapes of the Weibul (or similar) curves to fit the data which is available. For example, to estimate risk from established (rather than early) infections, or to model a scenario in which reliable tests are unavailable, one could modify the maximum detectability to below 1 (and similarly for specificity).

The code base developed as part of this work allows many aspects of risk modelling (and other applications) to be changed in a modular fashion - for example, the functions used to describe detectability or infectiousness curves could be adjusted or replaced without rewriting the rest of the code. In theory all the components might be substituted, in which case the existing code base could guide the user regarding which decisions must be made in order to sensibly use the general framework described in Chapter 2. In such cases the decisions themselves - e.g. how to model detectability, specifying mean delays between various events, and implementing inter-subject variability - would be made by the user.

Stellenbosch University https://scholar.sun.ac.za

Blood banking beyond the WHO guidelines

In severely resource-constrained contexts, limited funds are available for the screening of donor-derived material. In addition, the prevalence and incidence of transfusion-transmitted infections are often significantly higher in resource-limited settings [101, 102, 103]. Approved screening tests exist for a wide number of transfusion-transmissible infections, and 'non-approved' (by local or international approval bodies) tests might also be used. Procedures for screening donor-derived material in such contexts should ideally be such that the negative impact of TTI's on public health is minimized - for example by maximising Quality Adjusted Life Years (QALY) with a fixed per-donation budget, or optimizing the QALY gained per currency spent. There is however little published literature on the extent of formal optimization in any blood banking systems, particularly in the developing world [80, 104, 105].

The risk of TTI's in blood transfusion services following WHO best-practice guidelines is low, even in high-prevalence, high-incidence settings [36, 2]. This raises some questions. Are we focussing our scientific efforts in the right direction? What is the state of transfusion safety in low-resource settings, and how do TTIs contribute to ongoing epidemics of established TTI's (e.g. HBV, HBC, HIV, syphilis, West-Nile virus) and outbreaks of emerging TTIs? What would the cost-effectiveness of improving transfusion safety be, and how well do these figures compete with existing public health spending in such settings?

The modelling approaches described in this work could be naturally extended in order to explore residual risk implications of new screening algorithms, incorporating whatever level of data is available for characterising the tests in question.

Modelling the introduction of a pathogen inactivation technology

Pathogen-inactivation methods, which reduce the levels of viable pathogens in a sample, thus reducing transmission potential of potentially present pathogens, have been developed for various blood products and are considered by some to be the future of blood safety [106]. Due to the wide scope of the term and the novelty of the technology, approval standards for pathogen inactivation methods are based on the somewhat arbitrary reduction factor of 10^6 [106]. Given

Stellenbosch University https://scholar.sun.ac.za

that this is several orders of magnitude beyond the necessary reduction for some TTI's, it has been suggested that lower levels of pathogen inactivation should be approved, for use in conjunction with NAT testing in some form. The semi-mechanistic model and extensions could be applied to this scenario using the existing code with well-chosen parametric inputs. For example, if it was assumed that viable virions were uniformly inactivated by a factor of 10^3, but that the RNA from these virions remained detectable, then the 'pool size' parameter could be set to a value of $\frac{1}{10^3}$. On the other hand, if experiments found variability between donors in the extent of inactivation, this could be incorporated into the distribution of per-virion infectiousness values [106, 107].

6.3 Conclusion

Modelling the performance of tests for disease and infection has proved valuable in a number of fields, including blood safety and infection dating. The approach taken in this work represents a conceptual shift from previous risk estimation methods. We found that methods without variability work well in most domains, but produce unrealistic results when the mean delay between detectability and infectiousness is small or negative. Additionally, it was demonstrated that, when inter- and intra-subject variability is considered, reducing the diagnostic delay of test algorithms has diminishing returns, since some risk remains even when the mean delay between infectiousness and detectability is zero or negative. On the other hand, when samples are pooled for a NAT assay, increasing pool sizes cause progressively smaller increases in the residual risk values which they imply. The framework described in this work, and used to generate these insights, with the accompanying code base, may be readily applied to fields such as infection dating (see Grebe et al. [16]), incidence estimation, and estimation of the residual risk of transfusion-transmitted infections.

Characterising test performance in terms of diagnostic sensitivity and specificity as face-value numbers obscures dynamics which can add value to the way both scientists and clinicians conceptualise test performance. The context in which test performance is considered determines how much detail is appropriate to include in a model, and the implications of nuanced approaches encom-

Stellenbosch University https://scholar.sun.ac.za

passed by the generalised framework should be explored in order to make responsible decisions about modelling efforts which rely on realistic descriptions of test performance.

Stellenbosch University https://scholar.sun.ac.za

References

[1] Michael P Busch. Transfusion-transmitted viral infections: building bridges to transfusion medicine to reduce risks and understand epidemiology and pathogenesis. *Transfusion*, 46(9):1624–1640, 2006.

[2] RY Dodd. Transfusion-transmitted infections: testing strategies and residual risk. *ISBT Science Series*, 9(1):1–5, 2014.

[3] Herbert A Perkins, Susan Samson, Jane Garner, Dean Echenberg, JR Allen, Morton Cowan, and Jay A Levy. Risk of AIDS for recipients of blood components from donors who subsequently developed AIDS. *Blood*, 1987.

[4] Roger Y Dodd. Transmissible Disease and Blood Transfusion. *Science*, 186(4169):1138–1140, 1974.

[5] Hans Vrielink and Henk W Reesink. Virus transmission by allogenous blood and blood components. *Bailliere's Clinical Anaesthesiology*, 11(2), 1997.

[6] Jon J van Rood, Frans HJ Claas, Anneke Brand, Marcel GJ Tilanus, and Cees van Kooten. Half a century of Dutch transplant immunology. *Immunology Letters*, 162(2):145–149, 2014.

[7] Daniele Prati. Transmission of hepatitis C virus by blood transfusions and other medical procedures: a global review. *Journal of Hepatology*, 45(4):607–616, 2006.

[8] Michael P Busch and Roger Y Dodd. NAT and blood safety: what is the paradigm? *Transfusion*, 40(10):1157–1160, 2000.

Stellenbosch University https://scholar.sun.ac.za

[9] Roger Y Dodd. The risk of transfusion-transmitted infection. *New England Journal of Medicine*, 327:419–421, 1992.

[10] George B Schreiber, Michael P Busch, Steven H Kleinman, and James J Korelitz. The risk of transfusion-transmitted viral infections. *New England Journal of Medicine*, 334(26):1685–1690, 1996.

[11] Steven Kleinman, Michael P Busch, James J Korelitz, and George B Schreiber. The incidence/window period model and its use to assess the risk of transfusion-transmitted human immunodeficiency virus and hepatitis C virus infection. *Transfusion Medicine Reviews*, 11(3):155–172, 1997.

[12] RY Dodd. Transmission of parasites by blood transfusion. *Vox Sanguinis*, 74(S2):161–163, 1998.

[13] Eberhard W Fiebig, David J Wright, Bhupat D Rawal, Patricia E Garrett, Richard T Schumacher, Lorraine Peddada, Charles Heldebrant, Richard Smith, Andrew Conrad, Steven H Kleinman, et al. Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection. *Aids*, 17(13):1871–1879, 2003.

[14] Jos Weusten, Marion Vermeulen, Harry van Drimmelen, and Nico Lelie. Refinement of a viral transmission risk model for blood donations in seroconversion window phase screened by nucleic acid testing in different pool sizes and repeat test algorithms. *Transfusion*, 51(1):203–215, 2011.

[15] Jos JAM Weusten, Harry AJ Van Drimmelen, and P Nico Lelie. Mathematic modeling of the risk of HBV, HCV, and HIV transmission by window-phase donations not detected by NAT. *Transfusion*, 42(5):537–548, 2002.

[16] Eduard Grebe, , Shelley N. Facente, Jeremy Bingham, Christopher D. Pilcher, Andrew Powrie, Jarryd Gerber, Gareth Priede, Trust Chibawara, Michael P. Busch, Gary Murphy, Reshma Kassanjee, and Alex Welte. Interpreting HIV diagnostic histories into infection time estimates: analyt-

Stellenbosch University https://scholar.sun.ac.za

ical framework and online tool. *BMC Infectious Diseases*, 19(1), October 2019.

[17] Samuel Alizon and Carsten Magnus. Modelling the course of an HIV infection: insights from ecology and evolution. *Viruses*, 4(10):1984–2013, 2012.

[18] Bernhard P Konrad, Darlene Taylor, Jessica M Conway, Gina S Ogilvie, and Daniel Coombs. On the duration of the period between exposure to HIV and detectable infection. *Epidemics*, 20:73–83, 2017.

[19] A Bouchnita, G Bocharov, A Meyerhans, and V Volpert. Towards a multi-scale model of acute hiv infection. 2017, 5 (1), 6, 2017.

[20] Michael P Busch, Simone A Glynn, Susan L Stramer, D Michael Strong, Sally Caglioti, David J Wright, Brandee Pappalardo, Steven H Kleinman, and NHLBI-REDS NAT Study Group. A new strategy for estimating risks of transfusion-transmitted viral infections based on rates of detection of recently infected donors. *Transfusion*, 45(2):254–264, 2005.

[21] Simone A Glynn, Steven H Kleinman, David J Wright, Michael P Busch, et al. International application of the incidence rate/window period model. *Transfusion*, 42(8):966–972, 2002.

[22] Kevin P Delaney, Debra L Hanson, Silvina Masciotra, Steven F Ethridge, Laura Wesolowski, and S Michele Owen. Time until emergence of HIV test reactivity following infection with HIV-1: implications for interpreting test results and retesting after exposure. *Clinical Infectious Diseases*, 64(1):53–59, 2016.

[23] R. Biassoni and A. Raso. *Quantitative Real-Time PCR: Methods and Protocols*. Methods in Molecular Biology. Springer New York, 2014.

[24] D. Liu. *Molecular Detection of Human Bacterial Pathogens*. Taylor & Francis, 2011.

[25] R.W. Maitta. *Immunologic Concepts in Transfusion Medicine*. Elsevier Health Sciences, 2019.

Stellenbosch University https://scholar.sun.ac.za

[26] J. Linden. *Blood Safety and Surveillance*. CRC Press, 2001.

[27] J.K. Actor. *Introductory Immunology: Basic Concepts for Interdisciplinary Applications*. Elsevier Science, 2014.

[28] World Health Organization. Expert Committee on Biological Standardization and World Health Organization. *WHO Expert Committee on Biological Standardization: Sixty-sixth Report*. Public Health Series. World Health Organization, 2016.

[29] Reshma Kassanjee, Christopher D Pilcher, Sheila M Keating, Shelley N Facente, Elaine McKinney, Matthew A Price, Jeffrey N Martin, Susan Little, Frederick M Hecht, Esper G Kallas, et al. Independent assessment of candidate HIV incidence assays on specimens in the CEPHIA repository. *AIDS*, 28(16):2439, 2014.

[30] Shelley N Facente, Michael P Busch, Eduard Grebe, Christopher D Pilcher, Alex Welte, Brian Rice, and Gary Murphy. Challenges to the performance of current HIV diagnostic assays and the need for centralized specimen arcHIVes: a review of the Consortium for the Evaluation and Performance of HIV Incidence Assays (CEPHIA) repository. *Gates Open Research*, 3, 2019.

[31] Myron S Cohen, Cynthia L Gay, Michael P Busch, and Frederick M Hecht. The detection of acute HIV infection. *The Journal of Infectious Diseases*, 202(Supplement_2):S270–S277, 2010.

[32] Grifols Diagnostics Solution, Licence 2032, San Diego, CA. *Procleix Ultrio Assay*, 502623 rev. a edition, 2012. Accessed: 2019-10-02.

[33] Silvina Masciotra, J Steven McDougal, Jane Feldman, Patrick Sprinkle, Laura Wesolowski, and S Michele Owen. Evaluation of an alternative HIV diagnostic algorithm using specimens from seroconversion panels and persons with established HIV infections. *Journal of Clinical Virology*, 52:S17–S22, 2011.

Stellenbosch University https://scholar.sun.ac.za

[34] Eleftherios Mylonakis, Maria Paliou, Michelle Lally, Timothy P Flanigan, and Josiah D Rich. Laboratory testing for infection with the human immunodeficiency virus: established and novel approaches. *The American Journal of Medicine*, 109(7):568–576, 2000.

[35] WHO Expert Committee on Biological Standardization et al. Guidelines on estimation of residual risk of HIV, HBV or HCV infections via cellular blood components and plasma. *Report Series*, 1004:34, 2017.

[36] Marion Vermeulen, Nico Lelie, Charl Coleman, Wendy Sykes, Genevieve Jacobs, Ronel Swanevelder, Michael Busch, Gert van Zyl, Eduard Grebe, Alex Welte, et al. Assessment of HIV transfusion transmission risk in South Africa: a 10-year analysis following implementation of individual donation nucleic acid amplification technology testing and donor demographics eligibility changes. *Transfusion*, 59(1):267–276, 2019.

[37] World Health Organization. *Screening Donated Blood for Transfusion-transmissible Infections: Recommendations*. Nonserial Publication Series. World Health Organization, 2010.

[38] S. Carr, N. Unwin, T. Pless-Mulloli, and E. Corporation. *An Introduction to Public Health and Epidemiology*. Open University Press, 2007.

[39] B. Rosner. *Fundamentals of Biostatistics*. Cengage Learning, 2015.

[40] Myron S Cohen, George M Shaw, Andrew J McMichael, and Barton F Haynes. Acute HIV-1 infection. *New England Journal of Medicine*, 364(20):1943–1954, 2011.

[41] Ruy M Ribeiro, Li Qin, Leslie L Chavez, Dongfeng Li, Steven G Self, and Alan S Perelson. Estimation of the initial viral growth rate and basic reproductive number during acute HIV-1 infection. *Journal of Virology*, 84(12):6096–6102, 2010.

[42] Alan S Perelson. Modelling viral and immune system dynamics. *Nature Reviews Immunology*, 2(1):28, 2002.

Stellenbosch University https://scholar.sun.ac.za

[43] J Michael Kilby. Human immunodeficiency virus pathogenesis: insights from studies of lymphoid cells and tissues. *Clinical Infectious Diseases*, 33(6):873–884, 2001.

[44] John Coffin and Ronald Swanstrom. HIV pathogenesis: dynamics and genetics of viral populations and infected cells. *Cold Spring Harbor Perspectives in Medicine*, 3(1):a012526, 2013.

[45] Guy Severin Mahiane, Rachid Ouifki, Hilmarie Brand, Wim Delva, and Alex Welte. A general HIV incidence inference scheme based on likelihood of individual level data and a population renewal equation. *PLoS One*, 7(9):e44377, 2012.

[46] Reshma Kassanjee, Thomas A McWalter, Till Bärnighausen, and Alex Welte. A new general biomarker-based incidence estimator. *Epidemiology*, 23(5):721, 2012.

[47] Michael P Busch, Christopher D Pilcher, Timothy D Mastro, John Kaldor, Gaby Vercauteren, William Rodriguez, Christine Rousseau, Thomas M Rehle, Alex Welte, Megan D Averill, et al. Beyond detuning: 10 years of progress and new challenges in the development and application of assays for HIV incidence estimation. *Aids*, 24(18):2763–2771, 2010.

[48] Michael J Sweeting, Daniela De Angelis, John Parry, and Barbara Suligoi. Estimating the distribution of the window period for recent HIV infections: a comparison of statistical methods. *Statistics in Medicine*, 29(30):3194–3202, 2010.

[49] Christopher D Pilcher, Travis C Porco, Shelley N Facente, Eduard Grebe, Kevin P Delaney, Silvina Masciotra, Reshma Kassanjee, Michael P Busch, Gary Murphy, S Michele Owen, et al. A generalizable method for estimating duration of HIV infections using clinical testing history and HIV test results. *Aids*, 33(7):1231–1240, 2019.

[50] Eduard Grebe, Michael P Busch, Edward P Notari, Roberta Bruhn, Claire Quiner, Daniel Hindes, Mars Stone, Sonia Bakkour, Hong Yang, Phillip

Stellenbosch University https://scholar.sun.ac.za

Williamson, et al. HIV incidence in US first-time blood donors and transfusion risk with a 12-month deferral for men who have sex with men. *Blood, The Journal of the American Society of Hematology*, 136(11):1359–1367, 2020.

[51] Anthon du P Heyns, Richard J Benjamin, JP Ronel Swanevelder, Megan E Laycock, Brandee L Pappalardo, Robert L Crookes, David J Wright, and Michael P Busch. Prevalence of HIV-1 in blood donations following implementation of a structured blood safety policy in South Africa. *Jama*, 295(5):519–526, 2006.

[52] Chalmers TC Grady GF. Risk of post-transfusion viral hepatitis. *New England Journal of Medicine*, 271:337–342, 1964.

[53] John H Walsh, Robert H Purcell, Andrew G Morrow, Robert M Chanock, and Paul J Schmidt. Posttransfusion hepatitis after open-heart operations: incidence after the administration of blood from commercial and volunteer donor populations. *Jama*, 211(2):261–265, 1970.

[54] Leonard B Seeff. The history of the "natural history" of hepatitis C (1968 - 2009). *Liver International*, 29:89–99, 2009.

[55] Ronald E Domen. Paid-versus-volunteer blood donation in the United States: a historical review. *Transfusion Medicine Reviews*, 9(1):53, 1995.

[56] Baruch S Blumberg and Harvey J Alter. A new antigen in leukemia sera. *Jama*, 191(7):541–546, 1965.

[57] David J Gocke, Harry B Greenberg, and Neil B Kavey. Correlation of Australia antigen with posttransfusion hepatitis. *Jama*, 212(5):877–879, 1970.

[58] David J Gocke, HarryB Greenberg, and NeilB Kavey. Hepatitis antigen: Detection of infectious blood donors. *The Lancet*, 294(7614):248–249, 1969.

[59] Harvey J Alter, Paul V Holland, Robert H Purcell, Jerrold J Lander, Stephen M Feinstone, Andrew G Morrow, and Paul J Schmidt. Posttransfusion hepatitis after exclusion of commercial and hepatitis-B antigen-positive donors. *Annals of Internal Medicine*, 77(5):691–699, 1972.

Stellenbosch University https://scholar.sun.ac.za

[60] SH Kleinman, JC Niland, SP Azen, EA Operskalski, LH Barbosa, AI Chernoff, VM Edwards, BA Lenes, GJ Marshall, GJ Nemo, et al. Prevalence of antibodies to human immunodeficiency virus type 1 among blood donors prior to screening, the transfusion safety study/NHLBI donor repository. *Transfusion*, 29(7):572–580, 1989.

[61] MP Busch, MJ Young, SM Samson, JW Mosley, JW Ward, HA Perkins, and Transfusion Safety Study Group. Risk of human immunodeficiency virus (HIV) transmission by blood transfusions before the implementation of HIV-1 antibody screening. *Transfusion*, 31(1):4–11, 1991.

[62] Herbert A Perkins, Susan Samson, and Michael P Busch. How well has self-exclusion worked? *Transfusion*, 28(6):601–602, 1988.

[63] Jennifer C Learmont, Rosemary S Phillips, and Ian J Bickerton. The value of lookback to understanding blood-borne infectious diseases: The New South Wales' HIV experience. *Transfusion Medicine Reviews*, 16(4):315–324, 2002.

[64] Mindy Goldman and Gwendoline Spurll. Hepatitis C lookback. *Current Opinion in Hematology*, 7(6):392–396, 2000.

[65] M John Gill, Gwyneth Meyers, and Amin Rajwani. Use of blood donation history of people with HIV infection to identify recipients at risk. *CMAJ: Canadian Medical Association Journal*, 151(8):1147, 1994.

[66] T Zeiler, V Kretschmer, and W Sibrowski. A retrospective study of the practice of" look-back" procedures, on the incidence of HIV-1/2-positive blood donors and the risk of transfusion-associated HIV infection in public-community blood banks in Germany. *Infusionstherapie und Transfusionsmedizin*, 21(6):362, 1994.

[67] Michaël Chassé, Lauralyn McIntyre, Shane W English, Alan Tinmouth, Greg Knoll, Dianna Wolfe, Kumanan Wilson, Nadine Shehata, Alan Forster, Carl van Walraven, et al. Effect of blood donor characteristics on transfusion outcomes: a systematic review and meta-analysis. *Transfusion Medicine Reviews*, 30(2):69–80, 2016.

Stellenbosch University https://scholar.sun.ac.za

[68] Michael P Busch and Simone A Glynn. Use of blood-donor and transfusion-recipient biospecimen repositories to address emerging blood-safety concerns and advance infectious disease research: the national heart, lung, and blood institute biologic specimen repository, 2009.

[69] Rieko Sobata, Naoya Shinohara, Chieko Matsumoto, Shigeharu Uchida, Shigeru Igarashi, Satoru Hino, Masahiro Satake, and Kenji Tadokoro. First report of human immunodeficiency virus transmission via a blood donation that tested negative by 20-minipool nucleic acid amplification in Japan. *Transfusion*, 54(9):2361–2362, 2014.

[70] Marie-Claude Boily, Rebecca F Baggaley, Lei Wang, Benoit Masse, Richard G White, Richard J Hayes, and Michel Alary. Heterosexual risk of HIV-1 infection per sexual act: systematic review and meta-analysis of observational studies. *The Lancet Infectious Diseases*, 9(2):118–129, 2009.

[71] Pragna Patel, Craig B Borkowf, John T Brooks, Arielle Lasry, Amy Lansky, and Jonathan Mermin. Estimating per-act HIV transmission risk: a systematic review. *AIDS*, 28(10):1509, 2014.

[72] Rebecca F Baggaley, Richard G White, and Marie-Claude Boily. HIV transmission risk through anal intercourse: systematic review, meta-analysis and implications for HIV prevention. *International Journal of Epidemiology*, 39(4):1048–1063, 04 2010.

[73] Steven H Kleinman and Michael P Busch. The risks of transfusion-transmitted infection: direct estimation and mathematical modelling. *Best Practice & Research Clinical Haematology*, 13(4):631–649, 2000.

[74] Eve M Lackritz, Glen A Satten, John Aberle-Grasse, Roger Y Dodd, Vincent P Raimondi, Robert S Janssen, W Frank Lewis, Edward P Notari, and Lyle R Petersen. Estimated risk of transmission of the human immunodeficiency virus by screened blood in the United States. *New England Journal of Medicine*, 333(26):1721–1725, 1995.

[75] Roberta Bruhn, Nico Lelie, Brian Custer, Michael Busch, Steven Kleinman, and International NAT Study Group. Prevalence of human im-

Stellenbosch University https://scholar.sun.ac.za

munodeficiency virus RNA and antibody in first-time, lapsed, and repeat blood donations across five international regions and relative efficacy of alternative screening scenarios. *Transfusion*, 53(10pt2):2399–2412, 2013.

[76] Manuel Alvarez, Salvador Oyonarte, PM Rodriguez, and JM Hernandez. Estimated risk of transfusion-transmitted viral infections in Spain. *Transfusion*, 42(8):994–998, 2002.

[77] Josiane Pillonel, Syria Laperche, Christine Saura, Jean-Claude Desenclos, Anne-Marie Couroucé, and Transfusion-Transmissible Agents Working Group of the French Society of Blood Transfusion. Trends in residual risk of transfusion-transmitted viral infections in France between 1992 and 2000. *Transfusion*, 42(8):980–988, 2002.

[78] Claudio Velati, Luisa Romanò, Lorella Baruffi, Marco Pappalettera, Vittorio Carreri, and Alessandro R Zanetti. Residual risk of transfusion-transmitted HCV and HIV infections by antibody-screened blood in Italy. *Transfusion*, 42(8):989–993, 2002.

[79] Douglas R Bish, Ebru K Bish, Ryan S Xie, and Susan L Stramer. Going beyond "same-for-all" testing of infectious agents in donated blood. *IIE Transactions*, 46(11):1147–1168, 2014.

[80] Ebru K Bish, Prasanna K Ragavan, Douglas R Bish, Anthony D Slonim, and Susan L Stramer. A probabilistic method for the estimation of residual risk in donated blood. *Biostatistics*, 15(4):620–635, 2014.

[81] Anthony D Slonim, Ebru K Bish, and Ryan S Xie. Red blood cell transfusion safety: probabilistic risk assessment and cost/benefits of risk reduction strategies. *Annals of Operations Research*, 221(1):377–406, 2014.

[82] Sudha Jayaraman, Zaid Chalabi, Pablo Perel, Carla Guerriero, and Ian Roberts. The risk of transfusion-transmitted infections in sub-Saharan Africa. *Transfusion*, 50(2):433–442, 2010.

[83] Package Insert. Procleix® Ultrio™ Assay. *IN0167EN, Rev*, 1, 2004.

Stellenbosch University https://scholar.sun.ac.za

[84] P Nico Lelie, Harry AJ Van Drimmelen, H Theo M Cuypers, Susan J Best, Susan L Stramer, Catherine Hyland, Jean-Pierre Allain, Pierre Monchar-mont, Christine Defer, Micha Nübling, et al. Sensitivity of HCV RNA and HIV RNA blood screening assays. *Transfusion*, 42(5):527–536, 2002.

[85] Azzedine Assal, Valérie Barlet, Marie Deschaseaux, Isabelle Dupont, Pierre Gallian, Cathy Guitton, P Morel, Bernard David, and Philippe De Micco. Comparison of the analytical and operational performance of two viral nucleic acid test blood screening systems: Procleix Tigris and cobas s 201. *Transfusion*, 49(2):289–300, 2009.

[86] AS De Vos, RW Lieshout-Krikke, E Slot, EA Cator, and MP Janssen. A novel approach to detect test-seeking behaviour in the blood donor pop-ulation: making the invisible visible. *Vox Sanguinis*, 111(3):274–280, 2016.

[87] Center for Biologics Evaluation and Research. Complete List of DSA for Infectious Agents and HIV Diagnostic Assays. Accessed: 2018-07-01.

[88] Cecilia Noecker, Krista Schaefer, Kelly Zaccheo, Yiding Yang, Judy Day, and Vitaly V Ganusov. Simple mathematical models do not accurately predict early SIV dynamics. *Viruses*, 7(3):1189–1217, 2015.

[89] Steven H Kleinman, Nico Lelie, and Michael P Busch. Infectivity of hu-man immunodeficiency virus-1, hepatitis C virus, and hepatitis B virus and risk of transmission by transfusion. *Transfusion*, 49(11):2454–2489, 2009.

[90] AJ Leslie, KJ Pfafferott, P Chetty, R Draenert, MM Addo, M Feeney, Y Tang, EC Holmes, T Allen, JG Prado, et al. HIV evolution: CTL escape mutation and reversion after transmission. *Nature Medicine*, 10(3):282–289, 2004.

[91] José M Cuevas, Ron Geller, Raquel Garijo, José López-Aldeguer, and Rafael Sanjuán. Extremely high mutation rate of HIV-1 in vivo. *PLoS Biol*, 13(9):e1002251, 2015.

Stellenbosch University https://scholar.sun.ac.za

[92] Kevin K Ariën, Guido Vanham, and Eric J Arts. Is HIV-1 evolving to a less virulent form in humans? *Nature Reviews Microbiology*, 5(2):141–151, 2007.

[93] B Custer, C Quiner, R Haaland, A Martin, M Stone, R Reik, WR Steele, D Kessler, PC Williamson, SA Anderson, AE Williams, HF Raymond, W McFarland, WT Robinson, S Glick, K Sey, CD Melton, SA Glynn, SL Stramer, and MP Busch. HIV antiretroviral therapy and prevention use in US blood donors: a new blood safety concern. *Blood, The Journal of the American Society of Hematology*, 136(11):1351–1358, 2020.

[94] S.D. Pennypacker, H.D. Knoble, C.E. Antle, and L.V. Maden. A flexible model for studying plant disease progression. *Phytopathology*, 70:232–235, 1980.

[95] Yohann Foucher, Eve Mathieu, Philippe Saint-Pierre, Jean-François Durand, and Jean-Pierre Daurès. A semi-Markov model based on generalized Weibull distribution with an illustration for HIV disease. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 47(6):825–833, 2005.

[96] Zhen Wang, Chris T Bauch, Samit Bhattacharyya, Alberto d'Onofrio, Piero Manfredi, Matjaž Perc, Nicola Perra, Marcel Salathé, and Dawei Zhao. Statistical physics of vaccination. *Physics Reports*, 664:1–113, 2016.

[97] Andrew L Vizard, Garry A Anderson, and Robin B Gasser. Determination of the optimum cut-off value of a diagnostic test. *Preventive Veterinary Medicine*, 10(1-2):137–143, 1990.

[98] Karline Soetaert. *rootSolve: Nonlinear root finding, equilibrium and steady-state analysis of ordinary differential equations* , 2009. R package 1.6.

[99] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020.

[100] Andri Signorell et mult. al. *DescTools: Tools for Descriptive Statistics*, 2020. R package version 0.99.37.

Stellenbosch University https://scholar.sun.ac.za

[101] Angela Ugwu, Dalha Gwarzo, Theresa Nwagha, Aisha Gwarzo, and Andreas Greinacher. Transfusion in limited infrastructure locations–where to go decades after safe blood initiative by World Health Organization? *ISBT Science Series*, 15(1):118–125, 2020.

[102] Richard J Wright and Jeffrey SA Stringer. Rapid testing strategies for HIV-1 serodiagnosis in high-prevalence African settings. *American Journal of Preventive Medicine*, 27(1):42–48, 2004.

[103] Evan M Bloch, Eric A Gehrie, Paul M Ness, Jeremy Sugarman, and Aaron AAR Tobian. Blood transfusion safety in low-resourced countries: aspiring to a higher standard, 2020.

[104] Shiguang Xie. *Optimal Allocation of Resources for Screening of Donated Blood*. PhD thesis, Virginia Tech, 2011.

[105] Shaul K Bar-Lev, Onno Boxma, Igor Kleiner, David Perry, and Wolfgang Stadje. Recycled incomplete identification procedures for blood screening. *European Journal of Operational Research*, 259(1):330–343, 2017.

[106] SL Stramer, RY Dodd, and CY Chiu. Advances in testing technology to ensure transfusion safety–NAT and beyond. *ISBT Science Series*, 10(S1):55–64, 2015.

[107] Soraya Amar El Dusouqui, Marion C Lanteri, Rudolf Schwabe, Anja Grzesiczek, Anne North, Nina Mufti, Yassongui Mamadou Sekongo, John Pitman, and Claude Tayou Tagny. Progress towards an appropriate pathogen reduction technology for whole blood in Africa. *ISBT Science Series*, 15(1):151–163, 2020.

[108] Robert S Galen and S Raymond Gambino. *Beyond normality: the predictive value and efficiency of medical diagnoses*. Wiley New York, 1975.

[109] Cecilia Noecker, Krista Schaefer, Kelly Zaccheo, Yiding Yang, Judy Day, and Vitaly Ganusov. Simple mathematical models do not accurately predict early SIV dynamics. *Viruses*, 7(3):1189–1217, 2015.

Stellenbosch University https://scholar.sun.ac.za

[110] D Domanović. Assessing fusion-transmitted emerging infections. *ISBT Science Series*, 11(S1):68–75, 2016.

[111] B Custer. Assessing fusion-transmitted emerging infectious diseases. *ISBT Science Series*, 11(S2):79–85, 2016.

[112] M. Koffler. Epidemiology of hepatitis. *The Lancet*, 286(7417):834 – 835, 1965. Originally published as Volume 2, Issue 7417.

[113] George F Grady, Thomas C Chalmers, and Boston Inter-Hospital Liver Group. Risk of post-transfusion viral hepatitis. *New England Journal of Medicine*, 271(7):337–342, 1964.

[114] MP Busch. Insights into the epidemiology, natural history and pathogenesis of hepatitis C virus infection from studies of infected donors and blood product recipients. *Transfusion Clinique et Biologique*, 8(3):200–206, 2001.

[115] Paul B Beeson. Jaundice occurring one to four months after transfusion of blood or plasma: report of seven cases. *Journal of the American Medical Association*, 121(17):1332–1334, 1943.

[116] SL Stramer. Current perspectives in transfusion-transmitted infectious diseases: emerging and re-emerging infections. *ISBT Science Series*, 9(1):30–36, 2014.

[117] G Murphy, CD Pilcher, SM Keating, R Kassanjee, SN Facente, A Welte, E Grebe, K Marson, MP Busch, P Dailey, JM Garcia-Calleja, K Marsh, S Ongarello, J Osborn, and N Parkin. Moving towards a reliable HIV incidence test–current status, resources available, future directions and challenges ahead. *Epidemiology & Infection*, 145(5):925–941, 2017.

[118] Eduard Grebe, Alex Welte, Jake Hall, Sheila M Keating, Shelley N Facente, Kara Marson, Jeffrey N Martin, Susan J Little, Matthew A Price, Esper G Kallas, MP Busch, CD Pilcher, and G Murphy. Infection Staging and Incidence Surveillance Applications of High Dynamic Range Diagnostic Immuno-Assay Platforms. *Journal of Acquired Immune Deficiency Syndromes (1999)*, 76(5):547–555, 2017.

Stellenbosch University https://scholar.sun.ac.za

[119] Amelia Fisher, Selina Wallis, Oliver Hassall, Russell Martin, and Imelda Bates. Collaborations on blood transfusion research in sub-Saharan Africa: who, what and where. *Vox Sanguinis*, 115(3):221–232, 2020.

[120] Virginie Supervie, Jean-Paul Viard, Dominique Costagliola, and Romulus Breban. Heterosexual risk of HIV transmission per sexual act under combined antiretroviral therapy: systematic review and bayesian modeling. *Clinical Infectious Diseases*, 59(1):115–122, 2014.

[121] Debdatta Basu and Rajendra Kulkarni. Overview of blood components and their preparation. *Indian Journal of Anaesthesia*, 58(5):529, 2014.

[122] Rachita Sood, Rachel R Yorlets, Nakul P Raykar, Remya Menon, Hemant Shah, and Nobhojit Roy. The global surgery blood drought: frontline provider data on barriers and solutions in Bihar, India. *Global Health Action*, 12(1):1599541, 2019.

[123] Dorothy Kyeyune-Byabazaire and Heather A Hume. Towards a safe and sufficient blood supply in Sub-Saharan Africa. *ISBT Science Series*, 14(1):104–113, 2019.

[124] A Weimer, CT Tagny, JB Tapko, C Gouws, AAR Tobian, Paul Michael Ness, and EM Bloch. Blood transfusion safety in sub-Saharan Africa: a literature review of changes and challenges in the 21st century. *Transfusion*, 59(1):412–427, 2019.

[125] Brian Custer, Shimian Zou, Simone A Glynn, Julie Makani, Claude Tayou Tagny, Magdy El Ekiaby, Ester C Sabino, Nabajyoti Choudhury, Diana Teo, Kenrad Nelson, Emmanuel Peprah, LeShawndra Price, and Michael M Engelgau. Addressing gaps in international blood availability and transfusion safety in low-and middle-income countries: a NHLBI workshop. *Transfusion*, 58(5):1307–1317, 2018.

[126] R Rossenkhan, M Rolland, JPL Labuschagne, RC Ferreira, CA Magaret, LN Carpp, Iv Matsen, Y Huang, EE Rudnicki, Y Zhang, N Ndabambi, M Logan, T Holzman, MR Abrahams, C Anthony, S Tovanabutra, C Warth, G Botha, D Matten, S Nitayaphan, H Kibuuka, FK Sawe,

Stellenbosch University https://scholar.sun.ac.za

D Chopera, LA Eller, S Travers, ML Robb, C Williamson, PB Gilbert, and PT Edlefsen. Combining Viral Genetics and Statistical Modeling to Improve HIV-1 Time-of-Infection Estimation towards Enhanced Vaccine Efficacy Assessment. *Viruses*, 11(7):607, 2019.

[127] Shelley N Facente, Michael P Busch, Eduard Grebe, Christopher D Pilcher, Alex Welte, Brian Rice, and Gary Murphy. *Assays (CEPHIA) repository [version 1]*.

[128] Roger Y Dodd. Germs, gels and genomes. A personal recollection of 30 years in blood safety testing. In *Transmissible Diseases and Blood Transfusion*, pages 3–20. Springer, 2002.

[129] Roger Y Dodd. Current safety of the blood supply in the United States. *International Journal of Hematology*, 80(4):301–305, 2004.

[130] Emilia Vynnycky and Richard White. *An introduction to infectious disease modelling*. OUP Oxford, 2010.

[131] Harvey J Alter, Susan L Stramer, and Roger Y Dodd. Emerging infectious diseases that threaten the blood supply. In *Seminars in hematology*, volume 44, pages 32–41. Elsevier, 2007.

[132] Shimian Zou, Karen Fujii, Stephanie Johnson, Bryan Spencer, Nicole Washington, Edward Notari Iv, Fatemeh Musavi, Bruce Newman, Ritchard Cable, Jorge Rios, KL Hillyer, CD Hillyer, and RY Dodd. Prevalence of selected viral infections among blood donors deferred for potential risk to blood safety. *Transfusion*, 46(11):1997–2003, 2006.

[133] Shimian Zou, Anne F Eder, Fatemeh Musavi, Edward P Notari Iv, Chyang T Fang, Roger Y Dodd, and ARCNET Study Group. Implementation of the Uniform Donor History Questionnaire across the American Red Cross Blood Services: increased deferral among repeat presenters but no measurable impact on blood safety. *Transfusion*, 47(11):1990–1998, 2007.

[134] Susan L Stramer, F Blaine Hollinger, Louis M Katz, Steven Kleinman, Peyton S Metzel, Kay R Gregory, and Roger Y Dodd. Emerging infectious

Stellenbosch University https://scholar.sun.ac.za

disease agents and their potential threat to transfusion safety. *Transfusion*, 49:1S–29S, 2009.

[135] Susan L Stramer, Roger Y Dodd, and AABB Transfusion-Transmitted Diseases Emerging Infectious Diseases Subgroup. Transfusion-transmitted emerging infectious diseases: 30 years of challenges and progress. *Transfusion*, 53(10pt2):2375–2383, 2013.

[136] Simone A Glynn, Michael P Busch, Roger Y Dodd, Louis M Katz, Susan L Stramer, Harvey G Klein, Graham Simmons, Steven H Kleinman, Susan B Shurin, and 2011 NHLBI Emerging Infectious Disease Task Force convened November 7. Emerging infectious agents and the nation's blood supply: responding to potential threats in the 21st century. *Transfusion*, 53(2):438–454, 2013.

[137] Susan L Stramer, Edward P Notari, David E Krysztof, and Roger Y Dodd. Hepatitis B virus testing by minipool nucleic acid testing: does it improve blood safety? *Transfusion*, 53(10pt2):2449–2458, 2013.

[138] CR Seed. Screening and confirmatory testing strategies for the major transfusion-transmissible viral infections. *ISBT Science Series*, 9(1):6–13, 2014.

[139] Prashant Pandey, Divya Setya, Roli Srivastava, and Mukesh K Singh. A prospective, observational study for optimization of antibody screening in pretransfusion compatibility testing. *Immunohematology*, 36(1):19–28, 2020.

[140] Herbert A Perkins and Michael P Busch. Transfusion-associated infections: 50 years of relentless challenges and remarkable progress. *Transfusion*, 50(10):2080–2099, 2010.

[141] Jean-Pierre Allain and Helen Lee. Rapid tests for detection of viral markers in blood transfusion. *Expert Review of Molecular Diagnostics*, 5(1):31–41, 2005.

[142] Muazzam Nasrullah, Laura G Wesolowski, William A Meyer III, S Michele Owen, Silvina Masciotra, Craig Vorwald, William J Becker, and

Stellenbosch University https://scholar.sun.ac.za

Bernard M Branson. Performance of a fourth-generation HIV screening assay and an alternative HIV diagnostic testing algorithm. *AIDS*, 27(5):731, 2013.

[143] Jean-Pierre Allain, Susan L Stramer, ABF Carneiro-Proietti, ML Martins, SN Lopes Da Silva, M Ribeiro, FA Proietti, and Henk W Reesink. Transfusion-transmitted infectious diseases. *Biologicals*, 37(2):71–77, 2009.

[144] Cristina R Pruett, Marion Vermeulen, Pete Zacharias, Charlotte Ingram, Claude Tayou Tagny, and Evan M Bloch. The use of rapid diagnostic tests for transfusion infectious screening in Africa: a literature review. *Transfusion Medicine Reviews*, 29(1):35–44, 2015.

[145] S Michele Owen, C Yang, T Spira, CY Ou, CP Pau, BS Parekh, D Candal, D Kuehl, MS Kennedy, D Rudolph, W Luo, N Delatorre, S Masciotra, ML Kalish, T Cowart, F amd Barnett, R Lal, and JS McDougal. Alternative algorithms for human immunodeficiency virus infection diagnosis using tests that are licensed in the United States. *Journal of Clinical Microbiology*, 46(5):1588–1595, 2008.

[146] Iain B Gosbell, Veronica C Hoad, Claire E Styles, June Lee, and Clive R Seed. Undetectable does not equal untransmittable for HIV and blood transfusion. *Vox Sanguinis*, 114(6):628–630, 2019.

[147] Angelo R Margaritis, Stewart M Brown, Clive R Seed, Philip Kiely, Bruno D'Agostino, and Anthony J Keller. Comparison of two automated nucleic acid testing systems for simultaneous detection of human immunodeficiency virus and hepatitis C virus RNA and hepatitis B virus DNA. *Transfusion*, 47(10):1783–1793, 2007.

[148] BN Muringani. What is the risk of contracting mycobacterium tuberculosis form donated blood? *Rev Infect Dis*, 5:216–226, 2017.

[149] Jan Graffelman and Fred van Eeuwijk. Calibration of multivariate scatter plots for exploratory analysis of relations within and between sets of variables in genomic research. *Biometrical Journal*, 47(6):863–879, 2005.

Stellenbosch University https://scholar.sun.ac.za

[150] Ryanne W Lieshout-Krikke, Dragoslav Domanovic, Wim de Kort, Wolfgang Mayr, Giancarlo M Liumbruno, Simonetta Pupella, Johann Kurz, Folke Knutson, Sheila MacLennan, and Gilles Folléa. Selection strategies for newly registered blood donors in European countries. *Blood Transfusion*, 15(6):495, 2017.

[151] World Health Organization. *Blood Donor Selection: Guidelines on Assessing Donor Suitability for Blood Donation*. NCBI Bookshelf. World Health Organization, 2013.